The Optimal Taxation of Business Owners

Thomas Phelan[*]
Federal Reserve Bank of Cleveland

March 2, 2023

Abstract

Business owners in the United States are disproportionately represented among the wealthy and are exposed to substantial idiosyncratic risk. Further, recent evidence indicates that business income primarily reflects returns to the human capital of the owner. Motivated by these facts, this paper characterizes stationary efficient allocations and signaling equilibria with linear taxes when business income depends on innate ability, luck, and the past effort of the owner. I first show that in stationary efficient allocations, more productive entrepreneurs typically bear more risk and the distributions of consumption and firm size are approximately Pareto, with the tail of the latter typically thicker than that of the former. I then consider allocations in which entrepreneurs may save in a risk-free bond and signal a willingness to exert continued effort by retaining a fraction of their wealth in their firm. The optimal linear taxation policy calls for separate treatment of firm profits, interest, and wealth. The tax on profits plays a redistributive role, the tax on interest affects the incentives to retain equity and exert effort, and the tax on wealth affects the degree of consumption smoothing over time.

Keywords: Optimal taxation, moral hazard, optimal contracting.

JEL Codes: D61, D63, E62.

1   Introduction

It is well-known that the distributions of income and wealth in the United States are skewed to the right and have become more unequal in recent decades. What is perhaps less well-known is that much of this increase is due to the growing importance of business income. Smith et al. [2019] and Cooper et al. [2016] 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. Further, Smith et al. [2019] also show that business income depends on the owners' active participation by documenting that the death of an owner on average leads to profits falling by more than half.[1]

The appropriate policy response to these facts naturally depends on their underlying causes, and in particular, whether business income reflects innate ability, current effort, or past investments. For instance, if income were risk-free and business income solely represented the return on savings, then the results of Chamley [1986] and Judd [1985] imply that business income ought not to be taxed at all, at least in the long run. If instead business income depended solely on exogenous productivity and the current choice of hours worked as in Mirrlees [1971], then as Diamond and Saez [2011] show, the top marginal tax rate may well be over 70 percent. However, both of these approaches miss something important about the above facts. If business income were solely the return to savings then the owner ought not to matter, and if productivity were unaffected by policy and owner actions it is not clear why its distribution should vary across time. Further, neither of the above views provides guidance on whether we ought to levy different taxes on profits, interest income, or capital gains.

Motivated by these considerations, this paper characterizes stationary efficient allocations and signaling equilibria with linear taxes in an environment in which business income is risky, grows over time, and depends on the past (unobserved) effort exerted by the owner. I first adopt a mechanism design approach and consider the problem of a planner who may choose any allocation which satisfies the constraints imposed by asymmetric information. I then use the insights obtained from this abstract framework to sharply characterize optimal linear taxes in an environment in which entrepreneurs signal a willingness to exert unobservable effort by promising to retain a fraction of their wealth in their firm.

In order to characterize constrained-efficient allocations, I first alter the principal-agent model of Sannikov [2008] to incorporate persistent effects of effort on output. I assume

---

[1]See page 1878 of Smith et al. [2019].

that the output of a firm at any moment is publicly observable and independent of current choices, but that the growth of its output over time is stochastic and depends on the owner's effort. To induce the owner to exert high effort, high realizations of output growth must be followed by either high consumption or leisure, leading to imperfect risk-sharing and inequality in all efficient allocations. For standard preferences, it becomes increasingly expensive to motivate agents with histories of high growth. When the expected return to effort is fixed over time as in Sannikov [2008], the principal therefore rewards rich agents with leisure (instead of only consumption) and such agents bear little to no risk. In contrast, the dependence of growth on effort in this paper implies that the return to effort is increasing in firm size, which has important implications for both the dynamics of risk-sharing and its distribution across income levels. Although it is still true that the cost of motivating agents rises with their consumption, in the efficient allocation of this paper agents typically obtain high consumption precisely because their productivity rose, and so the benefits of inducing effort rise along with the costs.

After characterizing the optimal contract, I show how the problem of a planner facing an economy with entrepreneurs and workers decomposes into a series of problems identical in form to a principal-agent problem. The persistent effect of effort has important implications for the implied stationary distributions of consumption and firm size, where the latter is measured either by output or employment. In contrast with Sannikov [2008], in which rich agents bear no risk and consumption is bounded above, the distribution of consumption in the stationary efficient allocations of this paper is approximately Pareto in the right tail. Further, when I restrict effort to be fixed over time, the stationary distributions of both consumption and firm size are double Pareto with tail parameters that can be sharply characterized in closed-form and which in general differ from one another. These "restricted-action" allocations, the optimal allocations attainable with constant effort, are both simple to characterize and provide intuition for the tail of the efficient allocation because the welfare loss of recommending the highest effort indefinitely vanishes as productivity rises.

Following this characterization I consider the implications for taxation policy. Prescriptions for taxes unavoidably depend on the degree of risk-sharing present in private markets and contracts that agents are capable of signing. At one extreme, if agents may write contracts of arbitrary complexity with financial intermediaries, then the optimal policy calls only for lump-sum transfers between agents at birth, because the government has no advantage over the private sector at overcoming agency frictions. However, since the assumption of such contracts may be unrealistic, to complement the above I characterize the optimal

3

linear taxes on profits, capital income and wealth in an environment in which agents may save in a risk-free bond and trade shares in their firms.

I consider equilibria of a signaling type, in which each entrepreneur announces at the initial date a fraction of their wealth that they will retain in their business in perpetuity, and at subsequent periods make consumption-savings and (unobservable) effort choices. The price per share of the firm depends on the beliefs of outside investors over the owner's effort, and the equilibrium notion imposes the requirement that these beliefs are true in equilibrium, given the incentives of the owner. Although the class of linear taxes does not exactly implement the efficient allocation, this taxation exercise draws upon the characterization of constant effort allocations to obtain novel insights in an environment with heterogeneous agents and endogenous productivity. Further, the welfare loss from adhering to linear taxes becomes arbitrarily small as the number of workers per entrepreneur grows large, and the optimal equilibrium allocation approaches the efficient allocation.

The dependence of productivity on unobserved effort implies an important effect of taxation policy in the presence of active equity markets: taxes on interest affect the incentives for outside investors to purchase shares in a business and hence affect the degree of risk-sharing in the private sector. Intuitively, the willingness of investors to purchase shares in a firm will depend on the incentives of the owner to exert continued effort to improve productivity, and hence on the fraction of their wealth invested in the firm. This in turn depends upon the outside option available to the owner, the return on risk-free savings, which is endogenous to taxation policy. These taxation results are reminiscent of Albanesi [2006], who shows that the need to provide incentives for effort can justify the double taxation of capital income in a two-period model in which physical capital is exogenously distributed equally across entrepreneurs. In contrast with Albanesi [2006], I allow firm size to be endogenous to tax policy, grow over time and to depend on the entire history of the owner's effort. The dependence of output on the owner's productivity ensures that unlike the case of physical capital, it is simply not feasible to reallocate all factors of production across agents. Consequently, in both the efficient and equilibrium allocations in this paper, even arbitrarily successful entrepreneurs remain exposed to the output of their firm and continue to exert effort to improve productivity, as only they are capable of doing so.

Under the above assumptions, the optimal stationary equilibrium with linear taxes calls for different taxes on profits, interest, and wealth, playing conceptually distinct roles. The redistributive role is played by the profits tax, as this effectively taxes the inelastic quantity of the model: the endowed ability of the firm owner. The profits tax reduces the

4

ex-ante value of the firm, but has no effect on the subsequent incentives to save or issue shares throughout the agent's life. In contrast, taxes on interest income and wealth alter the private return to effort in one's business and consumption smoothing, respectively. Further, whether or not the taxes on interest and wealth raise any revenue depends crucially on the nature of preferences. In particular, in the case of logarithmic utility, consumption is a martingale in the optimal competitive equilibrium and the taxes on interest income and wealth raise no revenue in the aggregate.

Related literature. A vast literature builds upon Mirrlees [1971] to derive optimal taxes in economies with asymmetric information. Scheuer [2014] considers entrepreneurs and firm formation and allows for pecuniary externalities across occupations and the possibility of occupation-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 [2017] study optimal taxation policy in the presence of "superstar" effects in the form of assortative matching between individuals and firms, and Ales and Sleet [2016] consider a similar environment in which the planner has an explicit concern for the welfare of shareholders. Although the above models incorporate business income, they are static and so no agent bears risk or is subject to a moral hazard problem.

The study of dynamic extensions of Mirrlees [1971] begins with Golosov et al. [2003], and is often referred to as the New Dynamic Public Finance. Most papers within this literature, such as Golosov et al. [2016] and Farhi and Werning [2013], focus on the properties of optimal taxes with exogenously evolving wages. Stantcheva [2017] considers an extension of this framework in which wages depend on both exogenous ability and the stock of human capital, but there is no moral hazard and no agent employs workers or trades shares in his business. Closer to the current paper are Albanesi [2006], Kapička and Neira [2019] and Best and Kleven [2012], who conduct optimal taxation exercises in two-period economies with hidden effort, and Makris and Pavan [2021], who conduct a similar analysis in an environment with learning-by-doing. However, these papers do not derive implications for long-run distributions of consumption or income, and because of their two-period nature cannot address how the risk borne by any agent depends on his history of productivity.[2]

An older literature within public finance, surveyed in Chari and Kehoe [1999] and often termed the "Ramsey" approach, characterizes optimal policy when the government is restricted to choosing linear taxes on consumption and income. Guvenen et al. [2019]

---

[2]Makris and Pavan [2021] calibrate their model to the average working life span in the US, but the productivity of agents only changes once and so the dynamics are identical to a two-period model.

work within this framework to analyze the merits of capital income and wealth taxes in an economy with heterogeneous entrepreneurs in which financial frictions inhibit the allocation of physical capital. In this setting capital income and wealth taxes place different burdens on productive and unproductive entrepreneurs and so affect allocative efficiency. In addition to the explicit modeling of moral hazard, the approach of the current paper differs from Guvenen et al. [2019] in allowing the productivity of a firm to depend on the past effort of the owner, and for agents to trade multiple assets (bonds and shares) in the decentralization.

Jones and Kim [2018] characterize competitive equilibria in an environment in which entrepreneurs are cannot save and the evolution of productivity is identical to that adopted in this paper. The combination of exponential growth and stochastic death can both generate a Pareto distribution for income and allow it to respond to policy. However, in contrast with Jones and Kim [2018], in the first part of this paper I do not impose a market structure and instead require only that allocations respect the restrictions imposed by asymmetric information. Further, by allowing business owners to both save in a risk-free bond and sell shares in their business, I can investigate how the portfolio decisions of owners depend on taxation policy. Ai et al. [2022], Shourideh [2013], and Phelan [2022] all show how a Pareto distribution of consumption may emerge in efficient allocations in environments with asymmetric information. The difference here is the nature of the agency problem: instead of allowing entrepreneurs to divert assets to consumption, in this paper entrepreneurs exert (unobservable) effort to improve firm productivity. Importantly, unlike all of the above papers, the equilibria in this paper do not possess the counterfactual property that the Pareto exponents for income and firm size must coincide. Although the focus of this paper is normative rather than positive, I emphasize that the optimal taxation exercise is conducted in an environment that can replicate this fact, and to the best of my knowledge, this distinguishes the current paper from the rest of the optimal taxation literature. Furthermore, the above papers assume that agents cannot issue equity, and so do not explore how the degree of private risk-sharing depends on taxation policy.

In this paper, output depends both on individual productivity and a shadow price of labor, and so the model allows me to study how changes in the number of entrepreneurs translate into changes in (efficient) inequality. I view this as complementary to the standard approach in the optimal taxation literature in which wages (or productivity) evolve exogenously, but current income is endogenous to taxation policy. Since I model a dynamic moral hazard problem with an active stock market, I simplify the ex-ante heterogeneity and leave the quantitative importance of these two margins for future work.

2 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 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.

Time is continuous and extends indefinitely. 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 $l \in [\underline{l}, 1]$, referred to as leisure, for some $\underline{l}$. In what follows it will sometimes be more natural to describe the economic intuition in terms of effort, which is defined as $e := 1 - l$. The agent has preferences over stochastic sequences of consumption and leisure given by

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

where the flow utility function is

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

for some $\gamma \geq 1$ and $\alpha \in (0, 1)$ and the $\gamma = 1$ case is interpreted as logarithmic utility, $u(c, l) = (1 - \alpha) \ln c + \alpha \ln l$. In what follows I will also write $\bar{\gamma} := 1 - (1 - \gamma)(1 - \alpha)$ for notational convenience. 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 variable θ referred to as his productivity. In this first section an agent of productivity θ inelastically produces a flow of θ units of output per time independent of his current action, and so productivity will be identified with output. In the general equilibrium setting of Section 3, output will depend on both the productivity of the owner and the wage of workers. The consumption and output produced by the agent are observable, while the productivity of the agent begins at the level $\theta_0 = 1$ and evolves stochastically over time in a manner depending on his 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 leisure $(l_t)_{t \geq 0}$ then productivity follows the law of motion,

$$d\theta_t = \mu_\theta(l_t)\theta_t dt + \sigma_\theta(l_t)\theta_t dZ_t \tag{3}$$

where $\sigma_\theta(l_t) = \sigma 1_{l_t < 1}$ for some $\sigma > 0$ and $\mu_\theta(l) := (\overline{\mu}_0 - (\overline{\mu}_0 - \overline{\mu}_1)l)1_{l_t < 1}$ for some $\overline{\mu}_0 > \overline{\mu}_1 \geq 0$. The specification of μ_θ and σ_θ indicates that the productivity of the agent stops evolving upon retirement. Finally, the principal is risk-neutral and discounts at the rate of time preference of the agent and so their preferences over output and consumption are

$$U^P(c,l) = \mathbb{E}\left[\int_0^\infty e^{-\rho t}[\theta_t 1_{l_1 < 1} - c_t]dt\right] \tag{4}$$

which expresses the assumption that output ceases upon retirement. The actions taken by the agent and principal may be arbitrary functions of the preceding history of output. The following definition formalizes this mathematically.

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

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

for all $t \geq 0$ almost surely.

Since the effort exerted by the agent is private information, the principal must restrict 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 when choosing a strategy, the agent understands how his actions change the probability of each history and weights them accordingly. For any strategy l I will write $\mathbb{E}^l$ for the associated expectation operator.[3] The utility of an agent confronted with an allocation (c, l^P) when adhering to l is then

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

and the definition of incentive compatibility is the following.

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

The principal's problem is then the following. It is indexed by the agent's initial productivity θ and the minimal level of utility U necessary for his participation.

---

[3]Formal definitions are given in Appendix A.1.

Definition 2.3. *Given utility U and productivity θ the problem of the principal is*

$$V(U,\theta) = \max_{(c,l)\in\mathcal{A}^{IC}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}[\theta_t 1_{l_t<1} - c_t]dt\right]$$

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

$$d\theta_t = \mu_\theta(l_t)\theta_t dt + \sigma_\theta(l_t)\theta_t dZ_t, \quad \theta_0 = \theta.$$

The independence of the shocks over time implies that the principal's problem is recursive in U and θ. When the allocation is deterministic, differentiating (5) with respect to time gives $\dot{U}_t = \rho(U_t - u(c_t,l_t))$. Standard arguments, reviewed in Appendix A.1, imply that in general incentive compatibility is equivalent to the change in utility equaling this deterministic term plus a stochastic term that may be sharply characterized as follows.

Proposition 2.1. *An allocation (c,l^P) is incentive compatible if and only there exists a process $(S_t)_{t\geq 0}$ such that $dU_t = \rho(U_t - u(c_t,l_t))dt + \rho\sigma S_t dZ_t^l$, where Z_t^l is standard Brownian motion and $S_t\mu_\theta(l_t^P) + u(c_t,l_t^P) \geq S_t\mu_\theta(l) + u(c_t,l)$ for all $l \in [\underline{l},1]$ and $t \geq 0$.*

Since the principal is risk-neutral and the agent risk-averse, the principal will choose S_t to be the smallest value satisfying the inequality in Proposition 2.1 for all $l \in [\underline{l},1]$ and $t \geq 0$. By differentiating with respect to the choice of effort (or equivalently, leisure) we see that the volatility of utility is then

$$\rho\sigma S_t = 1_{l<1}\frac{\rho\sigma\alpha}{(\overline{\mu}_0 - \overline{\mu}_1)l}(c^{1-\alpha}l^\alpha)^{1-\gamma} =: \sigma E(l)(1-\alpha)(c^{1-\alpha}l^\alpha)^{1-\gamma} \tag{7}$$

where the last equality defines the function E for brevity. Incentive compatibility is therefore equivalent to the elasticity of utility to output being sufficiently large to outweigh the benefits of shirking and is sometimes referred to as a "skin-in-the-game" constraint. This shows that utility must be more responsive to output when deviations are hard to detect or the benefits of deviation are large, and allows us to recast the principal's problem as an optimal control problem. Further, the assumptions on productivity growth together with the homotheticity of preferences allow for the following reduction to a single state variable.

Lemma 2.2. *For all U and θ we have $V(U,\theta) = V(U\theta^{\overline{\gamma}-1},1)\theta$ and the policy functions of the planner are functions of $U\theta^{\overline{\gamma}-1}$.*

Proof. Since incentive compatibility is unaffected if we scale consumption in every history

9

by the same scalar, for any $\lambda > 0$, the change-of-variable $(\theta_t, c_t) \mapsto (\lambda\overline{\theta}_t, \lambda\overline{c}_t)$ implies

$$V(U, \lambda\theta) = \lambda \max_{(\overline{c},l)\in\mathcal{A}^{IC}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}[\overline{\theta}_t 1_{l_t<1} - \overline{c}_t]dt\right]$$

$$U = \lambda^{1-\overline{\gamma}}\mathbb{E}\left[\int_0^\infty e^{-\rho t}u(\overline{c}_t, l_t)dt\right]$$

$$d\overline{\theta}_t = \mu_\theta(l_t)\overline{\theta}_t dt + \sigma_\theta(l_t)\overline{\theta}_t dZ_t, \quad \overline{\theta}_0 = \theta$$

which is exactly $\lambda V(U\lambda^{\overline{\gamma}-1}, \theta)$. $\qquad\square$

Unlike in Sannikov [2008], effort here has persistent effects on output and so utility is not sufficient to act as a state variable. However, Lemma 2.2 shows the principal's choices depend only on promised utility per unit of output, and so suggests the following definition.

Definition 2.4. *Given U and θ, define normalized utility and the normalized payoff function by $u := [(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}\theta^{-1}$ and $v(u) := V\left(u^{1-\overline{\gamma}}/(1-\gamma), 1\right)$, respectively.*

Lemma 2.2 shows that the optimal choices of the planner are functions only of normalized utility, which represents a kind of cost/benefit ratio for motivating the agent. Similar observations are made in Ai et al. [2022] and He [2009], where the agency problem involves the hidden diversion of resources rather than hidden effort. The normalized value function does not appear to possess a closed-form solution and numerical methods are necessary for its calculation. To gain some intuition for the structure of the optimal contract, it is instructive to characterize the value and policy functions associated with what I will term *restricted-action* allocations, in which the principal must recommend a fixed effort level for the entirety of an agent's life but is unconstrained in his choice of consumption. I focus on these because they may be characterized sharply in closed-form, and because the restricted-action value function for the highest effort level becomes an arbitrarily good approximation to the true value function as productivity becomes arbitrarily high.

To give a sense of why such allocations admit a simple characterization, note that if we rewrite consumption $c =: \overline{c}[(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}$ as a fraction of utility in consumption units, then Proposition 2.1 and (7) together imply

$$\frac{dU_t}{U_t} = \rho\left(1 - (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}\right)dt + \sigma E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}(1-\overline{\gamma})dZ_t.$$

For a fixed effort level the mean and volatility of utility growth depend only on $\overline{c}$ and not utility directly. Further, for these allocations the perturbation arguments of Golosov et al.

[2003] are applicable, and so the inverse Euler equation holds, which leads to the following simple characterization proved via a guess-and-verify approach in Appendix A.4.

Proposition 2.3. *For each $l \in [\underline{l}, 1]$ the restricted-action consumption function is $c_r(u; l) = \overline{c}_r(l)u := x(l)^{\frac{1}{1-\overline{\gamma}}} l^{-\frac{\alpha}{1-\alpha}} u$ and the restricted-action value function is*

$$v_r(u; l) = \frac{1_{l<1}}{\rho - \mu_\theta(l)} - \frac{\overline{c}_r(l)u}{\rho - \mu_c(l)}$$

where $\mu_c(l)$ denotes average consumption growth and $x(l)$ is the positive solution to the quadratic $0 = \sigma^2 E(l)^2(\overline{\gamma} - 1)(\overline{\gamma} - 1/2)x^2 + \rho x - \rho$. Further, $v_r'(u; l)$ is increasing in l.

There are two important points to be taken from Proposition 2.3. First, we obviously have $v(u) \geq v_r(u; l)$ for all $l \in [\underline{l}, 1]$, because imposing constancy of actions adds a constraint to the principal's problem, and so

$$\liminf_{u \to 0} v(u) \geq \lim_{u \to 0} v_r(u; \underline{l}) = \frac{1}{\rho - \mu_\theta(\underline{l})}. \tag{8}$$

Since $v(u) \leq (\rho - \mu_\theta(\underline{l}))^{-1}$ everywhere, it follows that the restricted-action function for the highest effort approximates the true value function near zero. Phrased differently, the loss to the principal per unit of output from adhering to this restricted-action allocation falls to zero as productivity rises. Intuitively, when confronted with an entrepreneur who is highly productive (relative to their promised utility), the principal will want her to work as hard as possible because the payoff from high effort is increasing in current productivity. It is always possible that productivity will later fall to a level where low effort is efficient, but this becomes increasing unlikely as the entrepreneur grows more productive, and the limit in (8) simply bears out this intuition. Also note that this characterization near zero depends on the preferences adopted and differs from Phelan and Townsend [1991] and Sannikov [2008], in which the agent will be retired at low levels of utility when utility is bounded below.

The second important consequence of Proposition 2.3 is that $v_r'(u; l)$ is increasing in leisure, which captures the intuitive fact that it is more expensive to provide an agent with a given level of utility when they exert higher effort. This provides insight into how effort varies with normalized utility and so allows us to build intuition for the dynamics of risk-bearing. To this end, consider the problem $v_r^*(u) := \max_{l \in [\underline{l}, 1]} v_r(u; l)$ of a principal who must fix effort throughout the agent's life, and note that Topkis' theorem implies that the associated leisure level is increasing in normalized utility. Such a "restricted" principal recommends high effort to individuals with low normalized utility and vice versa.

11

The expressions in Proposition 2.3 do not, by themselves, tell us how consumption and leisure evolve over time and in response to shocks. However, the closed-form expressions allow for a simple characterization of how risk-sharing varies over time.

Lemma 2.4. *In the restricted-action allocations associated with $l \in [\underline{l}, 1)$ the drift and diffusion of normalized utility satisfy $\mu_u - \sigma_u^2/2 = -\overline{\gamma}\sigma^2 E(l)^2 x(l)^2/2 - \mu_\theta(l) + \sigma_\theta(l)^2/2$ and high productivity shocks reduce normalized utility if and only if*

$$\rho < (1/\alpha - 1)(\overline{\mu}_0 - \overline{\mu}_1)l + \frac{\sigma^2}{2}(2\overline{\gamma} - 1)(\overline{\gamma} - 1). \tag{9}$$

Proof. Appendix A.2 applies Ito's lemma to show that

$$\mu_u - \frac{\sigma_u^2}{2} = \frac{\rho}{1 - \overline{\gamma}}(1 - (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}) + \frac{(\overline{\gamma} - 1)}{2}\sigma^2 E(l)^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma} - \mu_\theta(l) + \frac{\sigma_\theta^2(l)}{2} \tag{10}$$

which reduces to the claimed expression by Proposition 2.3. High productivity shocks reduce normalized utility when $\sigma_u \equiv \sigma(E(l)x(l) - 1) < 0$, or $E(l)x(l) < 1$. Since the quadratic in Proposition 2.3 is increasing for $x > 0$, this is equivalent to (9). $\qquad\square$

If a process X satisfies $dX_t = \mu_X X_t dt + \sigma_X X_t dZ_t$ for some constants μ_X and σ_X then it admits the representation $\ln(X_t/X_0) = (\mu_X - \sigma_X^2/2)t + \sigma_X Z_t$ almost surely for all $t \geq 0$. The long-run pathwise behavior is then determined not by the drift μ_X but instead by $\mu_X - \sigma_X^2/2$, which may be interpreted as a risk-adjusted growth rate. Lemma 2.4 shows that the risk-adjusted growth of normalized utility is the difference between two terms. The first, $-\overline{\gamma}\sigma^2 E(l)^2 x(l)^2/2$, represents a force common in economies with private information, in which the principal gradually immiserates the agent in order to relax future incentive constraints, while the second, $\mu_\theta(l) - \sigma_\theta(l)^2/2$, is the risk-adjusted growth in productivity.

The fact that effort affects growth rather than flow output implies that the dynamics of risk-sharing here are different from the seminal contributions of Phelan and Townsend [1991] and Sannikov [2008], in which the mapping from effort to output is fixed over time. In these models agents are typically retired at high levels of utility because they are too costly to motivate. 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 typically have high utility precisely because they experienced high productivity growth. After a series of favorable shocks, the cost of motivating the agent may have grown, but since actions affect the growth of output, so too has the benefit, and Lemma 2.2 shows that the principal cares solely about the ratio of these costs to benefits. Lemma 2.4 even

12

shows that in the restricted-action allocations, the dynamics of risk-sharing may be entirely reversed relative to models without persistent effects, in the sense that high shocks reduce normalized utility. Indeed, when inequality (9) is satisfied, higher productivity shocks lead the principal to recommend higher effort and more risk in the agent's utility. Further, for sufficiently high risk-aversion or uncertainty, or a sufficiently patient agent, this will be true regardless of the recommended leisure level.

Numerical illustration. I now compute some examples in order to illustrate the policy functions for consumption and leisure and the dynamics of risk and promised utility. For each variable of interest I will depict both the optimal restricted-action quantity alongside the true, efficient quantity. Note that since none of the above restricted-action value functions are globally optimal, the value function must be calculated numerically.[4] However, note that the optimal restricted-action value and policy functions coincide with the efficient counterparts for both low and high levels of normalized utility.

The purpose of this section is to illustrate the novel mechanisms in the model, and in particular the dynamics of utility in response to shocks. Guided in part by Lemma 2.4, I will therefore fix the parameters governing discounting, labor supply, and productivity growth at the standard values $(\rho, \alpha, \underline{\mu}_0, \overline{\mu}_1, \sigma, \underline{l}) = (0.06, 0.33, 0.06, 0.0, 0.2, 0.5)$ and consider the effect of varying the degree of risk-aversion γ. Note that the above specification for productivity amounts to minimum and maximum growth of 0 and 3 percent, respectively, and idiosyncratic volatility equal to the lower of the two values in Angeletos [2007]. Higher volatility or growth would only reinforce the points highlighted after Lemma 2.4.

I first set $\gamma = 2$. Figure 1 plots the policy functions for leisure and the coefficient of the consumption function (the $\overline{c}$ term in the above) together with the analogous optimal restricted-action functions. Note that in both the true optimal allocation and in the optimal restricted-action allocations, leisure is increasing in normalized utility and the coefficient of consumption is decreasing. This shows that as promised utility increases, the principal rewards the agent relatively more with leisure than consumption. These results for fixed productivity are qualitatively the same as those in Sannikov [2008].

However, as emphasized above, the persistent effects of effort on output in this environment has important implications for the evolution of normalized utility, and lead it to differ qualitatively from the model of Sannikov [2008]. Figure 2 plots the risk-adjusted growth in normalized utility and the volatility of normalized utility in both the efficient and optimal restricted-action cases. As noted subsequent to Lemma 2.4, the risk-adjusted

---

[4]Appendix F outlines the numerical method I use to solve the principal's problem.
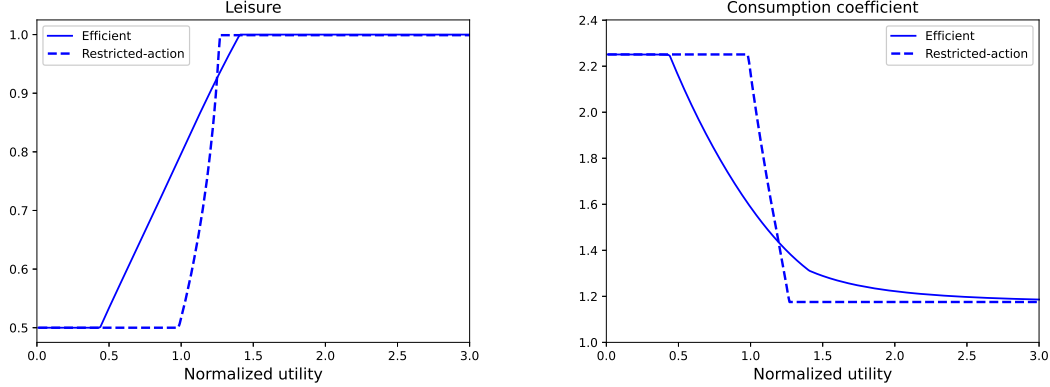
Figure 1: Efficient and restricted-action leisure and consumption.

growth in normalized utility is the difference between two terms, the first representing the risk-adjusted growth in the utility and the second the risk-adjusted growth in output. As expected, for low levels of output, the principal recommends high effort (low leisure), and the risk-adjusted growth in normalized utility is negative. Further, in Figure 2 the volatility of normalized utility is everywhere negative in both the efficient and the restricted-action allocations. When combined with the monotonicity of the leisure function, this implies that high realizations of productivity growth lead to lower normalized utility and therefore higher effort recommended by the principal. More succinctly, higher shocks lead to higher risk in future utility, and so as time goes on, the "rich" typically work harder and bear more risk than the "poor" in efficient allocations.

Figure 6 in Appendix C computes analogues of Figure 1 and Figure 2 when utility is logarithmic, which corresponds to the limit $\gamma \to 1$. In this case the risk-adjusted growth remains negative for small normalized utility, and the volatility is again everywhere negative, albeit closer to zero than in the $\gamma = 2$ case. I emphasize that the volatility of normalized utility is not guaranteed to be negative for all parameters.[5] However, the point of these examples is simply that the novel addition of persistent effects of effort can, in principle, completely reverse the relationship between output shocks and future effort, relative to the standard model of Sannikov [2008]. Further, in addition to being satisfied for the parameters given here, Lemma 2.4 shows that this reversal holds for all sufficiently high levels of risk.

---

[5]Indeed, it is easy to see that condition (9) in Lemma 2.4 can fail for, e.g., logarithmic utility and high discounting (high ρ) or higher preferences for leisure (higher α).
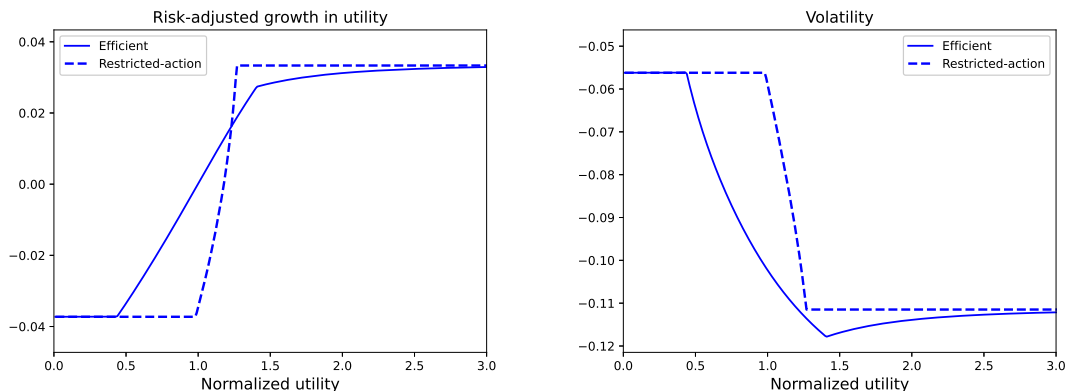
14

Figure 2: Efficient and restricted-action law of motion for utility.

In all of the above cases the efficient functions inherit the qualitative features of their restricted-action counterparts, but exhibit greater smoothness and vary less abruptly with normalized utility. Indeed, in numerical experimentations I have observed a kind of "bang-bang" solution for policy functions, with the lowest leisure level chosen at low levels of normalized utility, and the highest chosen at high levels.[6]

Before turning to the setting with a continuum of agents, it is instructive to compare the above optimal contract with those in other environments with hidden actions, and with the dynamics of income that obtain in economies with exogenously incomplete markets. This will also serve as a prelude as to why thick right tails of consumption and firm size emerge in this environment. Shourideh [2013] and Phelan [2022] both consider economies with hidden consumption. To preserve incentives for investment, the risk borne by entrepreneurs must scale with the benefits of diverting capital to consumption, and so rich agents with low marginal utility of consumption control high amounts of capital. With homothetic preferences, these forces cancel and the elasticity of consumption with respect to output is common across agents, and generates Pareto distributions in consumption and firm size. However, these two papers cannot generate the need for an active stock market in which share prices depend endogenously on the expected effort exerted by the owner.

It is also worth contrasting the lessons drawn from the current environment with the allocation in Jones and Kim [2018]. In their model entrepreneurs are assumed to be unable to save, and so consumption, income, and firm profits coincide. This differs sharply from

---

[6]This phenomenon appears to be more likely to occur when the degree of idiosyncratic risk is small.

the current paper, in which risk-sharing is restricted only by the need to provide incentives for effort. In particular, productivity growth affects consumption growth only insofar as it affects the return on effort and thereby the tightness of the incentive constraints. Consequently, although the stationary distributions of both consumption and firm size in the ensuing perpetual youth environment are both approximately Pareto, the right tail for firm size is typically thicker than that for consumption. Proposition 3.5 below formalizes this by showing that the distributions associated with the restricted-action allocations may be characterized in closed-form. While the tail of the firm size distribution depends only on the technology governing productivity growth and may be arbitrarily thick, the desire to smooth consumption over time places tight bounds on the tail of the ensuing distribution. Before proceeding to the general equilibrium context, I conclude this section by noting that if output is scaled by $\kappa \geq 0$, then the principal behaves as if confronted with an agent of productivity $\kappa\theta$. This has the following simple (but important) consequence.

Lemma 2.5. *For any parameter $\kappa > 0$, denote the normalized value function associated with this level of productivity by $v(u; \kappa)$. Then for all $\kappa, u \geq 0$, we have $v(u; \kappa) = \kappa v(u/\kappa; 1)$.*

Lemma 2.5 will prove useful in the general equilibrium setting when the productivity of entrepreneurs depends on an endogenous 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 shows how the problem of a 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. General equilibrium forces affect both the level of 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.

Time is again continuous and extends indefinitely. At any instant there is a continuum of agents alive in the economy with subjective discount factor ρ_S who die at rate ρ_D, so the preferences of agents continue to be of the form in (1) and (2) with discount rate $\rho := \rho_S + \rho_D$. A flow of ρ_D agents are born every unit of time so that the total population is fixed at unity. Every agent alive at the initial date is indexed by a scalar v identified with promised utility. For each agent there is an associated process $(Z_t^v)_{t \geq 0}$ distributed according

16

to standard Brownian motion and referred to as the noise process for the agent. The noise processes for subsequent generations will be indexed by their date of birth. These processes are independent of one another and so by a law of large numbers the ex-post distribution of shocks across agents will coincide with the ex-ante distribution faced by each agent.

Agents may be one of two types, entrepreneurs and workers, which are permanent and unobservable. A fraction $\eta_E \in [0, 1]$ are entrepreneurs and the remaining $1 - \eta_E$ are workers. Entrepreneurs are distinguished from workers by their ability to run businesses and have productivity that evolves as in Section 2, $d\theta_t^v = \mu_\theta(l_t)\theta_t^v dt + \sigma_\theta(l_t)\theta_t^v dZ_t^v$ with $\theta_0 = 1$. To ensure that output is finite in the stationary distribution I will also suppose that $\rho_D > \mu_\theta(l)$ for all leisure levels. In Section 2 the output of an entrepreneur coincided with his productivity. In contrast, in this section the output of an entrepreneur is a function of both their productivity and the effective labor assigned to him. If an entrepreneur of productivity θ is assigned L units of effective labor, then for a fixed $Z > 0$ and $\beta \in (0, 1)$ flow output of their firm is $F(\theta, L) = Z\theta^{1-\beta}L^\beta$ per unit of time. Workers inelastically supply an aggregate amont of $\overline{L}$ units of effective labor, while entrepreneurs oversee production and provide no effective labor. An allocation must specify the consumption, effort exerted, and labor assigned to every agent after every history. In the following $L_t^{v,\theta,i}$ and $L_t^{T,i}$ refer to the effective labor assigned to agents (which vanishes if $i = W$), and for agents not alive at the initial date, the superscript refers to birth date and the subscript to calendar time.

Definition 3.1. *Given a distribution Φ over utility, productivity, and types, an allocation consists of sequences $(c_t^{v,\theta,i}, l_t^{v,\theta,i}, L_t^{v,\theta,i})_{t\geq 0}$, $(v, \theta, i) \in \operatorname{supp}(\Phi)$ for the initial generation and $(c_t^{T,i}, l_t^{T,i}, L_t^{T,i})_{t\geq T\geq 0}, i = E, W$, for subsequent generations. An allocation satisfies promise-keeping if $U(c^{v,\theta,i}, l^{v,\theta,i}) = v$ for all $(v, \theta, i) \in \operatorname{supp}(\Phi)$.*

I will denote the set of all allocations by $\mathcal{A}$, and I will denote aggregate consumption, labor assignments, output, and (flow) utility by C_t, L_t, Y_t and U_t, respectively.[7]

Definition 3.2. *An allocation is resource feasible if it satisfies $C_t \leq Y_t$ and $L_t \leq \overline{L}$ for all $t \geq 0$. The set of such allocations will be denoted $\mathcal{A}^{RF}$.*

Incentive-compatibility requirements are of two kinds: an entrepreneur must be induced to reveal her type at birth and then to follow the effort recommendations of the planner. Incentive compatibility must, in principle, account for the possibility of double deviations, in which an agent misreports her type and then deviates from the recommended action.

---

[7]Formal definitions are contained in Appendix B.1.

However, such double deviations are irrelevant here, since an entrepreneur who claims to be a worker is subject to no further private information.

Definition 3.3. *The consumption and leisure allocations for a particular generation are incentive compatible if for all leisure strategies l', $U_E(c^E, l) \geq \max\left\{U_E(c^E, l'), U_W(c^W, 1)\right\}$. The set of incentive-compatible allocations will be denoted $\mathcal{A}^{IC}$ and the set of allocations that are incentive compatible and resource feasible is denoted $\mathcal{A}^{IF} := \mathcal{A}^{RF} \bigcap \mathcal{A}^{IC}$.*

For simplicity I will suppose that the planner only cares about workers and places a weight of $\alpha(T) = e^{-\rho_S T}$ on an agent born at time T, and write the utility attained by the planner over an allocation A by $U^P(A)$.[8]

Definition 3.4. *Given an initial distribution Φ over promised utility and types, the planner's problem is defined to be $V(\Phi) = \max_{A \in \mathcal{A}^{IF}} U^P(A)$.*

As in Farhi and Werning [2007] I focus on solutions in which the distributions of productivity and utility are constant over time. To this end, I first consider the simpler problem of a planner who may trade goods and labor intertemporally at the subjective rate of time preference. If we solve this *relaxed* problem and the implied distributions are constant over time, then we have solved the planner's problem beginning at that distribution. The relaxed problem has only two constraints and so the interdependence across agents is captured by two Lagrange multipliers, and so one need only maximize the components of the objective relevant to each generation in isolation. The only differences between this problem and the principal-agent problem analyzed in Section 2 are the requirement that the utility of an entrepreneur be sufficiently high to ensure truthful revelation, and the static task of assigning workers to entrepreneurs. At any moment the latter task depends solely on the shadow price of labor and productivity and solves $\max_{L \geq 0} \ Z\theta^{1-\beta}L^{\beta} - \lambda_L L$, the solution of which requires only elementary algebra and is summarized as follows.

Lemma 3.1. *Given the multiplier λ_L, labor assigned to an entrepreneur of productivity θ is $L(\theta) = [Z\beta/\lambda_L]^{\frac{1}{1-\beta}}\theta$. Flow output and output net of labor costs are $(1-\beta)^{-1}\overline{Z}(\lambda_L)\theta$ and $\overline{Z}(\lambda_L)\theta$, respectively, where $\overline{Z}(\lambda_L) := (1-\beta)Z^{\frac{1}{1-\beta}}(\beta/\lambda_L)^{\frac{\beta}{1-\beta}}$.*

Given a pair of multipliers $\lambda := (\lambda_R, \lambda_L)$, denote the associated stationary distribution over $\Omega' := \mathbb{R} \times \Theta \times \{E, W\}$ by Φ_λ. If this distribution satisfies the resource constraints

---

[8]This is equivalent to weighing flow utility independently of an agent's birth date. To aid the reader, elaboration of the social welfare function is given in Appendix B.1.

then the solution to the relaxed planner's problem with initial distribution Φ_λ amounts to adhering to the solutions to the generational planner's problem. I now combine the above with the optimal contract characterized earlier in partial equilibrium to infer properties of the long run distributions of consumption and productivity. Just as the homotheticity of preferences 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 attention to a one-dimensional distribution.

Definition 3.5. *Given a distribution Φ over the productivity and normalized utility, the summary measure is defined by $m(B) = \int_B \int_0^\infty \theta \Phi(\theta, u) d\theta du$ for any $B \subseteq [0, \infty)$.*

The homogeneity of the policy functions ensures that aggregate quantities may be expressed in terms of this summary measure. For instance, the average productivity and consumption of entrepreneurs is $\int_0^\infty m(u) 1_{l(u)<1} du$ and $\int_0^\infty c(u) m(u) du$, respectively. One implication of Lemma 3.1 is that changes in the shadow price of labor cause changes in the productivity of every entrepreneur, with payoffs remaining proportional to θ. When combined with Lemma 2.5, this implies that changes in resource constraints only affect normalized utility, with the subsequent policy functions identical to those found in the setting of Section 2. For any initial $\overline{u}$ denote the implied stationary density by $m_{\overline{u}}$, and note that the average productivity and consumption of entrepreneurs per unit of aggregate productivity in the stationary distribution may be written

$$M(\overline{u}) := \int_0^\infty m_{\overline{u}}(u) 1_{l(u)<1} du \qquad C(\overline{u}) := \int_0^\infty c(u) m_{\overline{u}}(u) du. \qquad (11)$$

Since workers have constant consumption throughout their lifetime, the associated aggregate consumption of workers is simply $(1 - \eta_E)\overline{Z}(\lambda_L)\overline{u}$. It remains to determine aggregate labor demand and consumption as functions of the multipliers and find conditions under which resources balance. Given an initial $\overline{u}$ the labor resource constraint reduces to

$$\overline{L} = (\text{ave. } \theta) \times (\text{no. entrepreneurs}) \times (\text{labor per } \theta)$$
$$= M(\overline{u}) \times \eta_E \times (Z\beta/\lambda_L)^{\frac{1}{1-\beta}}. \qquad (12)$$

Characterizing efficient allocations then reduces to solving a single nonlinear equation.

Proposition 3.2. *The stationary level of normalized utility is the solution to*

$$\frac{\eta_E}{1 - \beta} M(u) = (\eta_E C(u)/u + 1 - \eta_E)u \qquad (13)$$

with the associated level of utility in consumption units $(1 - \beta)Z(\eta_E M(u))^{-\beta}\overline{L}^\beta u$.

19

Proof. From Lemma 3.1 the output of each firm is $(1 - \beta)^{-1}\overline{Z}(\lambda_L)\theta$. Using (12) and integrating over all firms, aggregate output and consumption in the stationary distribution are $\eta_E(1-\beta)^{-1}\overline{Z}(\lambda_L)M(u)$ and $\overline{Z}(\lambda_L)(\eta_E C(u) + (1 - \eta_E)u)$, respectively. Dividing output and consumption by $\overline{Z}(\lambda_L)$ gives (13) and rearranging (12) gives utility. $\quad\square$

Proposition 3.2 shows 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. This allows for simple comparative statics.

Corollary 3.3. *The stationary level of normalized utility is increasing in η_E and β and independent of Z.*

Proof. Define $H(y) := M(y)(1 - \beta)^{-1} - C(y) - y(1 - \eta_E)/\eta_E$ and note that the stationary normalized utility is a root of H. All the claims follow from Topkis' theorem and the fact that H is increasing in η_E and β and independent of Z. $\quad\square$

The third claim in Corollary 3.3 may be viewed as a neutrality result, as it shows that changes in total factor productivity have no effect on the stationary value of normalized utility. Such changes therefore have no effect on inequality in consumption and firm size, as these are scaled for all agents after every history by the same amount.

Proposition 2.3 shows 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, and this is what is observed in all simulations. To extend the discussion following Lemma 2.4 and highlight how the efficient allocations of this paper differ from the exogenously incomplete markets model of Jones and Kim [2018], I now consider the stationary distributions associated with the restricted-action policy functions. This will also serve as a prelude to the results of Section 4, where it is shown that these distributions correspond to those that are implemented with linear taxes in a signaling equilibrium. The formulation of the stationary resource constraints is simpler and more transparent in this case. Recall from the discussion following Proposition 2.3 that v_r^* and l_r^* denote the value and policy functions of a principal who must recommend the same leisure level for the entirety of an agent's life. In this case consumption growth is characterized by constant mean

and volatility, and so for any leisure level l the stationary resource constraint becomes

$$Z\left(\frac{\rho_D\eta_E}{\rho_D - \mu_\theta(l)}\right)^{1-\beta}\overline{L}^\beta = \left(\frac{\rho_D\eta_E\overline{c}_{\mathrm{r}}(l)}{\rho_D - \mu_c(l)} + 1 - \eta_E\right)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}$$

where U denotes lifetime utility. Imposing the labor resource constraint once more and simplifying gives the following analogue of Proposition 3.2 and Corollary 3.3.

Proposition 3.4. *The optimal restricted-action leisure level is $l_{\mathrm{r}}^*(u_{\mathrm{r}})$, where u_{r} solves*

$$\frac{\rho_D\eta_E(1-\beta)^{-1}}{\rho_D - \mu_\theta(l_{\mathrm{r}}^*(u_{\mathrm{r}}))} = \left(\frac{\rho_D\eta_E\overline{c}_{\mathrm{r}}(l_{\mathrm{r}}^*(u_{\mathrm{r}}))}{\rho_D - \mu_c(l_{\mathrm{r}}^*(u_{\mathrm{r}}))} + 1 - \eta_E\right)u_{\mathrm{r}}.$$

The associated utility is $\overline{Z}u_{\mathrm{r}}$ where $\overline{Z} = (1-\beta)Z(\rho_D - \mu_\theta(l_{\mathrm{r}}^*(u_{\mathrm{r}})))^\beta(\overline{L}/[\rho_D\eta_E])^\beta$ and $l_{\mathrm{r}}^*(u_{\mathrm{r}})$ is increasing in both η_E and β.

The expressions in Proposition 2.3 then lead to the following characterization. A proof is contained in Appendix B.4.

Proposition 3.5. *For each $l \in [\underline{l}, 1]$ the stationary distributions of consumption and firm size associated with the restricted-action allocation with leisure l are both of the form $f(x) = Ax^{\alpha_X^+-1}1_{x\leq\overline{X}} + Bx^{\alpha_X^--1}1_{x>\overline{X}}$. The tail parameters for consumption are*

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

and the tail parameters for firm size (in output or employment) are

$$\alpha_\theta^\pm(l) = \mu_\theta(l)/\sigma^2 - 1/2 \pm \sqrt{(\mu_\theta(l)/\sigma^2 - 1/2)^2 + 2\rho_D/\sigma^2}.$$

Proposition 3.5 illustrates that the forces governing the distributions of firm size and consumption are very different from one another. For a given level of effort, the tail of the firm size distribution is mechanically determined by technological parameters.[9] In contrast, consumption growth depends on the technological parameters only insofar as the latter affect the tightness of the incentive constraints. Indeed, Proposition 3.5 shows that we have the uniform bound $\alpha_+(l) \leq -\overline{\gamma}$ for the tail parameter for consumption, independent of the nature of the agency problem. This places strong restrictions on the degree of inequality that can arise in efficient allocations, and shows that the implied distributions contrast

---

[9]A finite mean is assured if $\alpha_\theta^-(l) < -1$, which is implied by the assumption $\rho_D > \mu_\theta(l)$.

sharply with models with exogenously incomplete markets, such as Jones and Kim [2018], in which the tails of firm size and consumption are forced to coincide.

Numerical illustration. I now compute several examples depicting aggregate output and consumption in both the efficient and restricted-action allocations. In Figure 3, I first depict the average productivity and consumption of entrepreneurs, the functions M and C defined in (11), in both the efficient and optimal restricted-action allocations, where the parameters are fixed at those in Section 2. Notice that consumption of entrepreneurs need not be increasing in normalized utility, because as utility rises the planner may compensate the agents more with leisure than consumption. However, this need not lead to multiplicity of solutions to the resource constraints, because entrepreneurs are only one part of the population and workers' consumption is increasing in normalized utility. I earlier emphasized that the difference between the efficient allocation and the optimal restricted-action allocation vanishes as normalized utility falls. This observation is also reflected in the Figure 3, in which the dotted and solid lines coincide near the origin. Further, the solution to (13) necessarily tends to zero as the fraction of entrepreneurs become small, which reflects an intuitive point: as entrepreneurs become rarer (and hence more valuable to the economy), the efficient allocation becomes well-approximated by an allocation in which they work as hard as possible indefinitely, regardless of their history of productivity shocks.
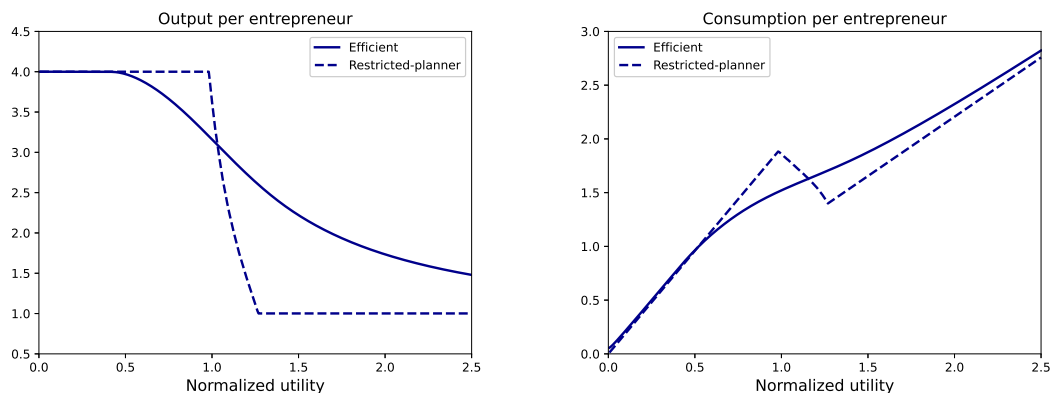

Figure 3: Output and consumption per entrepreneur and leisure.

I now plot the upper tails for consumption and firm size in both the efficient and restricted-action allocations associated with the highest effort. The sole parameters relevant to the general equilibrium setting in addition to those specified in Section 2 are β, η_E and

ρ_D, corresponding to the labor share, the mass of entrepreneurs, and the rate of death. I adopt the same parameters as those used to generate Figure 1, together with $\rho_D = 0.04$ and $\beta = 0.64$. If one views ρ_D as the rate with which one ceases to be an entrepreneur, then this amounts to assuming that agents run their business for an average of 25 years. I then consider a range of different values for η_E with midpoint 11.5 percent, which is the fraction of the population who are active business owners in the SCF as reported in Cagetti and De Nardi [2006]. Figure 4 plots the tails of the associated distributions of consumption and firm sizes (above the 90th percentile), alongside the analogous quantities for the restricted-action allocation associated with the highest effort.[10]
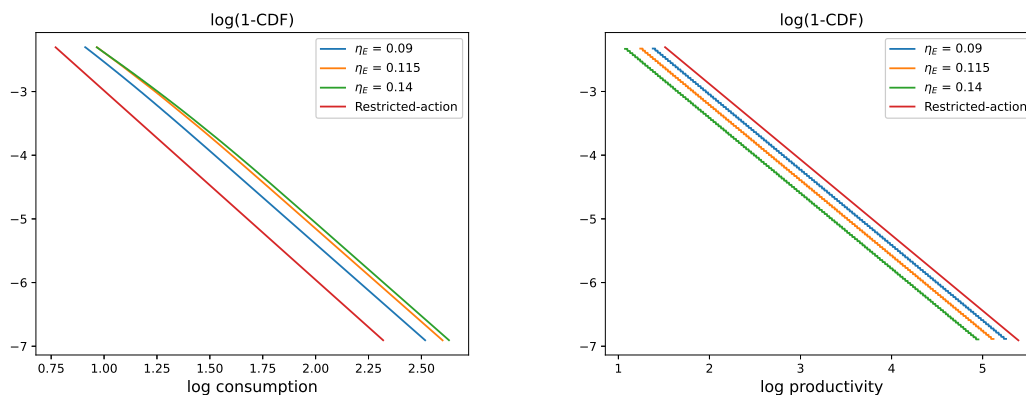

Figure 4: Tails for consumption and firm size.

As the mass of entrepreneurs varies, so too does the initial value of normalized utility solving equation (13). The lower this mass, the harder such entrepreneurs work and the more risk they bear throughout their lifetime. However, as suggested by the above discussion, ultimately the right tails of the distributions of consumption and firm size are dominated by entrepreneurs exerting the highest level of effort. Consequently, the shape of the right tail (i.e. the slope on a logarithmic scale) of the efficient allocations coincides with the corresponding shape of the restricted-action allocation for the highest effort.

---

[10]Appendix F outlines the numerical method used to produce these figures.

4 Optimal linear taxes

The foregoing analysis has focused on the forces that shape the long run efficient levels of inequality in an economy with repeated moral hazard, with all quantities implicitly specified by a social planner administering the direct mechanism. In this final section I discuss some implications for taxes. Prescriptions for taxes depend crucially on the contracts agents are capable of writing and enforcing. Indeed, at one extreme, by extending the results of Atkeson and Lucas [1992], when agents may write contracts of arbitrary complexity with intermediaries, the optimal policy consists only of lump-sum transfers between agents. However, the assumption of such a sector is quite strong, and so in this final section I fix a market structure and characterize the allocations attainable with constant linear taxes. I will draw on the characterization of the optimal constant-leisure allocations in Section 3 in order to understand how the decision to retain equity depends on taxation policy, and to highlight the distinct effects of taxes on profits, interest, capital gains, and wealth.

All agents may trade a risk-free government bond, and at the initial date, entrepreneurs decide on the fraction of their wealth that they will retain in their business in perpetuity, with the remaining shares sold to risk-neutral investors. The government levies constant linear taxes on firm profits and the owner's income and wealth, together with lump-sum transfers to workers. The choice of share retention of the entrepreneur is observable to all investors, who subsequently use this to form expectations of the entrepreneur's (unobservable) effort. The equilibrium notion will impose the requirement that these expectations be consistent with the choices of entrepreneurs, so that investors make zero expected profits.[11]

This market structure allows one to distinguish between firm profits, capital gains, and interest income, instead of absorbing them all into the category of "capital income", as would be the case if agents could only save in a risk-free bond. The allocations that arise with linear taxes turn out to coincide exactly with the restricted-action allocations characterized in Proposition 3.5, and the optimal taxes are simply chosen to ensure that the entrepreneurs exert the level of leisure given in Proposition 3.4. Consequently, even entrepreneurs with arbitrarily high wealth continue to exert effort, and the distributions of both firm size and wealth exhibit fat (Pareto) tails with indices that will, in general, differ from one another. The primary result of this section is that taxes on profits, interest income, and wealth

---

[11]This equilibrium notion differs from that adopted in earlier drafts, in which agents behaved competitively. Appendix E discusses these two solution concepts and shows that although they lead to different taxes, they raise the same amount of revenue whenever both allocations are well-defined.

play conceptually distinct and complementary roles, and that the tax on capital gains is redundant. The tax on profits achieves redistributive aims, the tax on interest income affects the incentives of owners to retain equity in their firm and exert effort, and the tax on wealth affects the degree of consumption smoothing over time.

I first describe the problem of an entrepreneur who has already committed to retaining a fraction of wealth in their firm, before turning to a discussion of this initial decision and of the appropriate equilibrium notion. The sole state variable for the entrepreneur will be financial wealth a_t, which is the sum of savings b_t and the value of firm equity $p_t x_t$, where x_t is the fraction of the firm owned by the agent and p_t is the (endogenous) share price,

$$a_t := b_t + p_t x_t. \tag{14}$$

Consistent with the assumptions of Section 3, I assume that entrepreneurs earn income only from business profits, interest and capital gains, and do not simultaneously receive a wage.[12] All agents may save in a risk-free government bond with before-tax return r_t. I also assume the existence of a competitive sector of insurance companies who trade actuarially fair annuity contracts with agents, promising a return ρ_D on the savings of agents in exchange for taking possession of these savings at death. The outside investors are risk-neutral, and so the price per share will equal the expected discounted value of firm profits, given the investors' expectations regarding the effort of the entrepreneur. The equilibrium notion will impose the requirement that these beliefs are correct, so that the outside investors break even in expectation. When investors believe that the effort exerted follows the process $\hat{l}$ and dividends are taxed according to the process τ_d, they value the firm at date t at

$$p_t = \mathbb{E}^{\hat{l}} \left[\int_t^\infty e^{-(r+\rho_D)(s-t)} (1 - \tau_{ds}) \overline{Z}(w_s) \theta_s ds \right],$$

where I remind the reader that $\overline{Z}$ is defined in Lemma 3.1 as output per unit of productivity net of labor costs (i.e. profits per θ_s), viewed as a function of the wage. I will restrict attention to stationary equilibria, in which case expectations and prices are constant and the price of the firm is simply given by the Gordon growth formula

$$p_t = \frac{(1 - \tau_d)\overline{Z}(w)\theta_t}{r + \rho_D - \mu_\theta(\hat{l})} \tag{15}$$

---

[12]This assumption does not drive the results that follow. The presence of wage income would complicate the choice of share retention but not the savings or effort choices, as consumption would simply be a constant fraction of total wealth, inclusive of the discounted value of labor income.

for all $t \geq 0$ almost surely. Note that because the share market is competitive, investors will make zero expected profits from every trade with entrepreneurs. When expectations $\hat{l}$ and taxes $(\tau_d, \tau_s, \tau_{cg}, \tau_a)$ on dividends, interest, capital gains, and wealth are constant over time, the entrepreneur's wealth evolves according to

$$da_t = [(1 - \tau_s)(r + \rho_D) - \tau_a - \bar{c}_t]a_t dt + \bar{\iota}_t a_t dR_t(l_t; \hat{l}) \tag{16}$$

where $\bar{c}_t$ and $\bar{\iota}_t$ denote the fraction of wealth consumed and invested, respectively, and

$$dR_t(l_t; \hat{l}) = (\tau_s(r + \rho_D) + (1 - \tau_{cg})\mu_\theta(l_t) - \mu_\theta(\hat{l}))dt + \sigma(1 - \tau_{cg})dZ_t \tag{17}$$

represents the excess return on retained equity. To understand the law of motion of wealth, note that the expected excess return on investing in the agent's business is the sum of after-tax dividends and capital gains minus the after-tax return on the risk-free bond, or

$$(1 - \tau_d)\overline{Z}\theta_t/p_t + (1 - \tau_{cg})\mu_\theta(l_t) - (1 - \tau_s)(r + \rho_D).$$

When combined with the price-earnings ratio given in (15), this gives the expression in (17) for the excess return on equity.[13] Notice that one implication of the expression (17) is that although the expected return is increasing in the effort of the entrepreneur, it is decreasing in the effort expected by investors, because high effort expectations imply a high price-earnings ratio, and hence low dividends per unit of equity retained.

Before turning to a formal definition of the entrepreneur's problem, there are several aspects of the law of motion of wealth in (16) that are worth discussing. First, it is important to note the distinction between $\hat{l}$ and l: the former is determined by the level of effort expected by investors while the latter is determined by the effort exerted by the agent. The two must coincide in equilibrium but only the latter is directly chosen by the entrepreneur, with the former jointly determined by the initial choice of share retention and the taxation policy of the government. Second, firm size does not appear in the law of motion of wealth, because with constant expectations, the price-dividend ratio is constant and hence so too is the expected return on retained equity. Third, the tax on firm profits and the wage also do not appear in the law of motion of wealth, because these only affect the value of the firm and hence the initial wealth of the agent, and not its evolution over time.

The fact that the restricted-action allocations are characterized by three constants implies that we should expect one degree of freedom in our choice of the four taxes. I therefore

---

[13]To aid the reader in the interpretation of the market structure and timing, a discrete-time formulation of the law of motion is contained in Appendix D.1.

assume that capital gains are not taxed, as this simplifies the expressions that follow, and I will also suppose that the government lends to the agents at $r = \rho_S$ and runs a primary surplus or deficit to balance its budget. In this case, the excess return on retained equity simplifies to

$$dR_t(l_t; \hat{l}) = (\tau_s\rho + \mu_\theta(l_t) - \mu_\theta(\hat{l}))dt + \sigma dZ_t \tag{18}$$

and so whenever the expected and actual leisure levels coincide, the expected excess return on equity is simply $\tau_s\rho$. Note that this excess return is increasing in the tax on interest because the latter reduces the outside option available to the entrepreneur, the risk-free bond, but does not affect the earnings-price ratio, which by (15) is always

$$(1 - \tau_d)\overline{Z}\theta_t/p_t = r + \rho_D - \mu_\theta(\hat{l}) = \rho - \mu_\theta(\hat{l})$$

regardless of taxes. The fact that the tax on interest appears in the excess return on equity but the tax on wealth does not hints at why they play conceptually distinct roles in the below implementation. The tax on interest alters the relative returns on the two assets, while the tax on wealth simply affects the growth rate of consumption.

I will now outline the entrepreneurs' problem and the equilibrium notion adopted in this setting. For clarity, I will split the individual problem in to two parts. First, I will characterize the optimal consumption and effort consistent with a given level of share retention (and hence expectations of investors). Second, I will impose rational expectations on the part of investors, and characterize the initial choice of share retention by the entrepreneur.[14]

Definition 4.1. *Given linear taxes* $\tau = (\tau_s, \tau_a)$ *and expectations governed by* $\hat{l}$, *the problem of an entrepreneur with wealth* a *who has pledged to retain* $\bar{\iota}$ *of her wealth in the firm is*

$$V(a; \bar{\iota}, \hat{l}) = \max_{\bar{c}, l} \mathbb{E}^l \left[\int_0^\infty \rho e^{-\rho t} u(\bar{c}_t a_t, l_t) dt \right]$$
$$da_t = [(1 - \tau_s)\rho - \tau_a - \bar{c}_t]a_t dt + \bar{\iota}_t a_t dR_t(l_t; \hat{l})$$
$$a_0 = a$$

where dR_t *is given by equation* (18).

Using a homogeneity argument almost identical to that given in Lemma 2.2, one may show that the optimal consumption is a constant fraction of wealth and that the optimal effort is constant in wealth.[15] I therefore denote by $\bar{c}^*(\bar{\iota}, \hat{l})$ and $l^*(\bar{\iota}, \hat{l})$ the optimal choices

---

[14]In the following I substitute $r = \rho_S$, so the pre-tax return on the bond is $\rho = \rho_S + \rho_D$.

[15]Explicit expressions for the policy functions are given in Lemma 4.1.

of consumption and leisure for the problem in Definition 4.1 as functions of the retained wealth share and expectations of investors. Lemma 4.2 below then shows that there is a unique solution to the equation $\hat{l} = l^*(\bar{\iota}, \hat{l})$, which I will denote by $\hat{l}(\bar{\iota})$ and which represent the leisure consistent with rational expectations. The problem of the entrepreneur at the initial date is then the following.

Definition 4.2. *Given a wage w and constant linear taxes $\tau = (\tau_d, \tau_s, \tau_a)$, the problem of an entrepreneur choosing the fraction of their initial wealth to retain is given by $\max_{\bar{\iota} \geq 0} V(\bar{a}_0(\bar{\iota}); \bar{\iota}, \hat{l}(\bar{\iota}))$, where V is given in Definition 4.1 and*

$$\bar{a}_0(\bar{\iota}) = \frac{(1 - \tau_d)\overline{Z}(w)\theta_0}{\rho - \mu_\theta(\hat{l}(\bar{\iota}))} \tag{19}$$

denotes the initial wealth of the owner as a function of $\bar{\iota}$.

The form of the problem in Definition 4.2 illustrates that the entrepreneur faces a simple tradeoff when considering the fraction of their wealth to retain in the firm at the initial date. On the one hand, retaining shares exposes them to idiosyncratic risk in wealth and hence consumption, which reduces their welfare. However, retaining shares also increases their subsequent incentives to exert effort, and so increases their initial wealth, because it increases the price outside investors are willing to pay per share at the initial date. Further, this tradeoff is endogenous to tax policy, because the excess return on equity depends on the tax on savings, and all investors rationally recognize this dependence.

Given the above formulation of the market structure and expectation formations, the equilibrium notion I adopt is of a standard type: individuals optimize, markets clear, and expectations are consistent with individual incentives.

Definition 4.3 (Signaling equilibrium). *Given constant linear taxes τ, a stationary signaling equilibrium consists of wages w, expectations $\hat{l}$, and policy functions $(\bar{c}, l) = (\bar{c}_t, l_t)_{t \geq 0}$ and $\bar{\iota}$ for consumption, leisure, and share retention, such that:*

- *The policy functions $(\bar{c}, l, \bar{\iota})$ solve the consumer problem given in Definition 4.1 and Definition 4.2, given expectations $\hat{l}$.*

- *The outside investors make zero expected profits, or $l_t = \hat{l}$ for all $t \geq 0$ almost surely.*

- *The markets for goods and labor clear every instant.*

- *The government budget constraint is satisfied.*

Prior to characterizing signaling equilibria, I wish to emphasize once more why the characterization of restricted-action allocations in Section 2 is useful for determining the optimal linear taxes. Indeed, the mechanism design approach adopted earlier may be appear to be superfluous given that in this section we restrict attention to linear taxes and make no mention of incentive compatibility. However, the key insight here is that for any level of share retention, the solution to the problem in Definition 4.1 with constant linear taxes implies constant effort by the entrepreneur. Given the planner's objective, these allocations are therefore weakly dominated by the restricted-action allocations, as the latter are, by definition, the optimal incentive compatible allocations with constant effort. Consequently, if we can implement the optimal restricted-action allocation as a signaling equilibrium, then we have found the optimal linear taxes, because the fact that entrepreneurs freely choose their consumption and (unobservable) leisure implies that the equilibrium allocation is incentive compatible. Further, Lemma 2.4 shows that the restricted-action allocations are completely characterized by the initial consumption level and the constant mean and volatility of consumption growth. It will therefore suffice to find linear taxes such that these quantities in the signaling equilibrium coincide with their counterparts in the optimal restricted-action allocations, because then the allocations will coincide almost surely at every date. The following algebra simply shows precisely how this is done.

I first characterize the expectations for effort consistent with a given choice of $\bar{\iota}$ and individual optimality, before turning to the initial choice of share retention. The proof of the following proceeds via a guess-and-verify approach and so is relegated to Appendix D.2.

Lemma 4.1. *Given a retained wealth share $\bar{\iota}$ and expectations $\hat{l}$, the optimal consumption and leisure choices are given by*

$$\bar{c}(\bar{\iota}, \hat{l}) = \frac{\rho}{\gamma} + \frac{1}{\gamma}\Big([\bar{\mu}_0\hat{l} + \tau_s\rho]\bar{\iota} - \bar{\gamma}\sigma^2\bar{\iota}^2/2 - \tau_a + (1 - \tau_s)\rho\Big)(\bar{\gamma} - 1)$$

and $l^*(\bar{\iota}, \hat{l})\bar{\mu}_0\bar{\iota}(1 - \alpha) = \alpha\bar{c}(\bar{\iota}, \hat{l})$ provided that $l^*(\bar{\iota}, \hat{l}) \in [\underline{l}, 1]$.

Lemma 4.1 allows expectations over leisure to be arbitrary. Note that because $\bar{\gamma} > 1$, the optimal consumption (and hence leisure) choice is increasing in the level expected by outside investors. As noted in the discussion following the expression for the return (17), a fall in the expected growth of output increases the expected return, and hence to an increase in consumption and leisure under the assumption $\bar{\gamma} > 1$. Further, Lemma 4.1 implies

$$\frac{\partial l^*}{\partial \hat{l}} = \frac{\alpha}{\bar{\mu}_0\bar{\iota}(1 - \alpha)}\frac{\partial c}{\partial \hat{l}} = \alpha(\gamma - 1)/\gamma < 1,$$

29

which ensures that there is a unique level of leisure at which expectations are consistent with individual optimality. The characterization of leisure, consumption and utility in any signaling equilibrium as a function of the initial level of share retention is then the following, the proof of which is contained in Appendix D.2.

Lemma 4.2. *For any $\bar{\iota}$ there is a unique solution $l(\bar{\iota})$ to $l = l^*(\bar{\iota}, l)$, the entrepreneur's consumption and leisure are $c(\bar{\iota}) = \rho\tilde{x}(\bar{\iota}; \tau)$ and $l(\bar{\iota}) = (\rho/\overline{\mu}_0)\alpha(1 - \alpha)^{-1}\tilde{x}(\bar{\iota}; \tau)\bar{\iota}^{-1}$, and their value function is a multiple of*

$$\tilde{x}(\bar{\iota}; \tau)^{-\overline{\gamma}}\bar{\iota}^{-\alpha(1-\overline{\gamma})}\frac{a^{1-\overline{\gamma}}}{1 - \overline{\gamma}} \tag{20}$$

where I define

$$\tilde{x}(\bar{\iota}; \tau) := \frac{1}{\overline{\gamma}}\left[1 + \left(\tau_s\bar{\iota} - \overline{\gamma}\sigma^2\bar{\iota}^2/[2\rho] - \tau_a/\rho + 1 - \tau_s\right)(\overline{\gamma} - 1)\right]. \tag{21}$$

The quadratic function $\tilde{x}$ defined in Lemma 4.2 does not appear to admit a simple economic interpretation. I simply defined it in the above manner so that consumption volatility may be written as $\sigma_c = \sigma E(l(\bar{\iota}))\tilde{x}(\bar{\iota}; \tau)$, which will ensure that the optimal taxes in what follows will be those which imply $\tilde{x}(\bar{\iota}; \tau) = x(l(\bar{\iota}))$ for the function x studied in the partial equilibrium setting of Section 2.[16]

Given the above analysis of the optimal choice of leisure and consumption consistent with rational expectations, it remains to determine the initial level of share retention. The wealth of the entrepreneur at the initial date is given by

$$a_0 = \frac{(1 - \tau_d)\overline{Z}(w)\theta_0}{\rho - \overline{\mu}_0(1 - \hat{l}(\bar{\iota}))}.$$

Using the above algebra, we have the following.

Lemma 4.3. *Given taxes τ_a and τ_s on wealth and interest, the problem of the entrepreneur when choosing the fraction of their wealth $\bar{\iota}$ to retain in their business is equivalent to*

$$\max_{\bar{\iota}} - \ln\left(\rho - \overline{\mu}_0 + \rho\alpha(1 - \alpha)^{-1}\tilde{x}(\bar{\iota}; \tau)\bar{\iota}^{-1}\right) + \frac{\overline{\gamma}}{\overline{\gamma} - 1}\ln\tilde{x}(\bar{\iota}; \tau) - \frac{\alpha}{1 - \alpha}\ln\bar{\iota} \tag{22}$$

In the case of log utility, the problem of the entrepreneur becomes

$$- \ln\left(\rho - \overline{\mu}_0 + \rho\alpha(1 - \alpha)^{-1}\bar{\iota}^{-1}\right) + \tau_s\bar{\iota} - \sigma^2\bar{\iota}^2/[2\rho] - \frac{\alpha}{1 - \alpha}\ln\bar{\iota}. \tag{23}$$

In all cases, the optimal choice of wealth retained is increasing in the tax on interest.

---

[16]Also note that $\tilde{x} \equiv 1$ identically in the case of logarithmic utility, and that $\tilde{x}$ is always concave in the share retention under the maintained assumption $\gamma \geq 1$.

Proof. The problem of the entrepreneur becomes

$$\max_{\bar{\iota}} \frac{1}{1-\gamma} \left(\rho - \overline{\mu}_0 + \rho\alpha(1-\alpha)^{-1}\tilde{x}(\bar{\iota};\tau)\bar{\iota}^{-1} \right)^{\overline{\gamma}-1} \tilde{x}(\bar{\iota};\tau)^{-\gamma}\bar{\iota}^{-\alpha(1-\gamma)} \tag{24}$$

Multiplying by $1-\gamma$, taking logs, and dividing by $1-\overline{\gamma}$ gives the expression in the statement, while the expression for logarthmic utility arises as we take the limit as $\gamma \to 1$. The fact that the optimal $\bar{\iota}$ is increasing in the tax on interest follows from Topkis' theorem. $\square$

As I will explain below, in certain special cases (like logarithmic utility) the above problem can be considerably simplified. For a given pair of taxes $\tau = (\tau_s, \tau_a)$ on interest and wealth, I will denote the associated choice of share retention of the entrepreneur by $\bar{\iota}(\tau)$. For arbitrary parameters and taxes, a sharp characterization of this choice is difficult, but the comparative statics and the conceptual roles played by each tax are most transparent and are the primary concern here. The final claim in Lemma 4.3 implies that the risk borne by the entrepreneur in any signaling equilibrium allocation is an increasing function of the tax on interest. Intuitively, an increase in this tax increases the excess return on retained equity, and therefore increase the incentive of the entrepreneur to expose themself to the performance of their firm.

Once again I recall that the optimal restricted-action allocations are completely characterized by three constants, specifying initial consumption and the mean and variance of consumption growth. For an initial level of wealth, in order to find the optimal linear taxation policy it will suffice to choose taxes on interest and wealth that ensure that the inverse Euler equation holds and that the volatility of consumption growth coincides with

$$\sigma_c^*(u_{\mathrm{r}}) := \sigma E(l_{\mathrm{r}}^*(u_{\mathrm{r}}))x(l_{\mathrm{r}}^*(u_{\mathrm{r}})) \tag{25}$$

where u_{r} is given in Proposition 3.4. One technical problem that can arise in this characterization is that although the optimal share retention choice in Lemma 4.3 is unambiguously increasing in the tax on interest, it might not be continuous (or even single-valued), because the objective faced by the entrepreneur is not always concave. It is therefore not clear that every interior value of the share retention can arise as we vary taxes. For now I will simply state this formally as an assumption and verify ex-post that it holds under weak conditions.

Assumption 4.1. *For each τ_a there exists τ_s such that $\sigma\bar{\iota}(\tau) = \sigma_c^*(u_{\mathrm{r}})$.*

To proceed further in the characterization of taxes, we use the fact that the inverse Euler equation holds in the restricted-action allocations. This imposes the following relationship between the taxes on wealth and interest.

Lemma 4.4. *Given the choice of share retention $\bar{\iota}$, the inverse Euler equation will hold in the signaling equilibrium if and only if taxes satisfy $\tau_a + \rho\tau_s(1 - \bar{\iota}) = -\overline{\gamma}(1 - \overline{\gamma})\sigma^2\bar{\iota}^2$.*

Proof. If μ_c and σ_c denote the mean and variance of consumption growth, then the inverse Euler equation is equivalent to $\mu_c - \sigma_c^2/2 = -\overline{\gamma}\sigma_c^2/2$. Using the law of motion of wealth, $da_t = [\rho - \bar{c} - \tau_a - \rho\tau_s(1 - \bar{\iota})]a_t dt + \sigma\bar{\iota}a_t dZ_t$, we then require

$$-\tau_a/\rho + 1 - \tau_s - \bar{c}/\rho + \tau_s\bar{\iota} = (1 - \overline{\gamma})\sigma^2\bar{\iota}^2/[2\rho].$$

Using the expression for $\bar{c}$ in Lemma 4.2, this last requirement successively simplifies to

$$1 - \tau_s - \tau_a/\rho + \overline{\gamma}\tau_s\bar{\iota} = \overline{\gamma}(1 - \overline{\gamma})\sigma^2\bar{\iota}^2/[2\rho] + 1 + \left(\tau_s\bar{\iota} - \overline{\gamma}\sigma^2\bar{\iota}^2/[2\rho]\right)(\overline{\gamma} - 1)$$

and $-\tau_a/\rho + 1 - \tau_s + \tau_s\bar{\iota} = 1 + \overline{\gamma}(1 - \overline{\gamma})\sigma^2\bar{\iota}^2/\rho$ and gives the desired expression. $\square$

Lemma 4.4 shows that for each interest tax there is a unique level of the wealth tax at which the inverse Euler equation holds. Lemma 4.3 and Lemma 4.4 imply that the characterization of the optimal taxes is quite elementary in this setting and may be thought of as consisting of two distinct steps. We first ensure that the efficient level of effort is implemented by varying the interest tax and using Lemma 4.3, and we then choose the wealth tax using Lemma 4.4. This is summarized as follows. I provide a sketch of the proof in the main text and relegate further details to Appendix D.3.

Proposition 4.5. *Under Assumption 4.1, the optimal linear taxation policy consists of choosing taxes on interest and assets satisfying the pair of equations*

$$\sigma\bar{\iota}(\tau) = \sigma_c^*(u_{\mathrm{r}})$$
$$\tau_a + \rho\tau_s(1 - \sigma_c^*(u_{\mathrm{r}})/\sigma) = -\overline{\gamma}(1 - \overline{\gamma})\sigma_c^*(u_{\mathrm{r}})^2$$

together with linear taxes on firm profits, lump-sum transfers to workers and government debt issuance such that all agents obtain the utility given in Proposition 3.4.

Proof sketch. The main claim that must be verified is that if taxes are chosen such that the entrepreneur bears consumption risk σ_c and the inverse Euler equation holds, then the entrepreneur will choose leisure l satisfying $\sigma_c = \sigma E(l)x(l)$. In other words, I need to check that if the taxes induce the entrepreneur to experience the "skin-in-the-game" associated with a particular leisure level in the restricted-action allocation, then the entrepreneur

actually chooses that level in the signaling equilibrium. Using Lemma 4.2, for any $\bar{\imath}$, the equilibrium consumption volatility is

$$\sigma_c = \sigma\bar{\imath} = \sigma E(l(\bar{\imath}))[1 + \left(\tau_s\sigma_c/\sigma - \bar{\gamma}\sigma_c^2/[2\rho] - \tau_a/\rho + 1 - \tau_s\right)(\bar{\gamma} - 1)]/\bar{\gamma}.$$

From Lemma 4.4, if taxes are chosen such that the inverse Euler equation holds, then

$$\sigma_c = \sigma E(l)\left[1 - (\bar{\gamma} - 1)(\bar{\gamma} - 1/2)\sigma_c^2/\rho\right],$$

while consumption volatility in the efficient allocation is $\sigma E(l)x(l)$. It follows that equilibrium and efficient consumption volatility satisfy the same quadratic, and so coincide. $\qquad\square$

The main insight of Proposition 4.5 is that the optimal linear taxation policy calls for separate taxes on wealth and capital income. The precise expressions for government debt, transfers, and the dividends tax are of secondary importance, as they are simply chosen such that the resource constraint holds and entrepreneurs are indifferent between revealing their type. I therefore relegate explicit expressions for these quantities to Appendix D.3. As emphasized above, the redistributive tool in this environment is the tax on firm profits, as this reduces the value of each entrepreneur's firm ex-ante. The taxes on capital income and wealth affect the incentives to retain equity and consumption smoothing. Further, whether or not these taxes on capital income and wealth actually raise any revenue turns out to depend crucially on the nature of preferences, as the following shows.

Corollary 4.6. *Under Assumption 4.1 the revenue per unit of wealth raised from the optimal choice of linear taxes on interest and wealth is $\bar{\gamma}(\bar{\gamma} - 1)\sigma_c^*(u_r)^2$. Further, if $\gamma > 1$ this quantity is increasing in the number of workers per entrepreneur.*

Proof. The revenue raised from the wealth and interest taxes is $\tau_a + \rho\tau_s(1 - \bar{\imath})$ and so the first claim follows from Lemma 4.4, while the second claim follows from Proposition 3.4. $\quad\square$

To gain intuition for Corollary 4.6, recall that the sign of the drift in consumption in the restricted action allocations coincides with the sign of $1 - \bar{\gamma}$. The drift in consumption is therefore negative precisely when revenue is being raised from the entrepreneurs.

Discussion and special cases. The optimal taxes in Proposition 4.5 are stated in terms of the share retention function $\bar{\imath}(\tau)$ defined in Lemma 4.3. However, this function is only defined implicitly and the corresponding taxes are only characterized under Assumption 4.1. In this final section I show that Assumption 4.1 is satisfied for several common

33

functional forms and that the above algebra simplifies substantially to yield intuitive expressions and further insights. Throughout this section I impose $\overline{\mu}_0 = \rho$, which is satisfied for the parameters considered in Section 2 and ensures that the problem (22) simplifies to

$$\max_{\overline{\iota} \geq 0} \frac{1 + \alpha(\gamma - 1)}{\overline{\gamma} - 1} \ln \tilde{x}(\overline{\iota}; \tau) + \left(\frac{1 - 2\alpha}{1 - \alpha} \right) \ln \overline{\iota}. \tag{26}$$

This objective is concave if $\alpha \leq 1/2$, in which case the first-order condition is

$$0 = \frac{[1 + \alpha(\gamma - 1)]}{(\gamma - 1)(1 - 2\alpha)} \tilde{x}'(\overline{\iota}; \tau)\overline{\iota} + \tilde{x}(\overline{\iota}; \tau). \tag{27}$$

There are at least two cases in which we obtain substantial simplification.[17] The first case is when utility is logarithmic. In this case, the consumption and leisure in Lemma 4.1 are independent of taxes and the function given in (21) is identically unity. Now the quadratic in (27) does not depend on the wealth tax, and the interest tax simplifies to

$$\tau_s = \frac{\sigma^2/\rho}{(1/\alpha - 1)l^*(u_{\mathrm{r}})} + (2 - 1/\alpha)l^*(u_{\mathrm{r}}). \tag{28}$$

In the above discussion, I emphasized the distinct conceptual roles played by each tax. One additional insight that emerges clearly in the logarithmic case is that in principle, the tax on interest can assume either sign. Indeed, the tax on interest is negative for sufficiently small volatility uniformly in leisure when $\alpha < 1/2$.

The second case in which the above characterization simplifies is when $\alpha = 1/2$.[18] In this case, the utility cost of high effort cancels with the higher initial wealth and the optimal share retention now simply maximizes the quadratic $\tilde{x}$ in equation (21) and implies that $\tau_s = \rho^{-1}\overline{\gamma}\sigma^2\overline{\iota}$, and so the optimal tax on interest is

$$\rho\tau_s = \overline{\gamma}\sigma\sigma_c^*(u_{\mathrm{r}}) \tag{29}$$

and the tax on wealth becomes

$$\tau_a = \overline{\gamma}^2\sigma_c^*(u_{\mathrm{r}})^2 - \overline{\gamma}\sigma\sigma_c^*(u_{\mathrm{r}}). \tag{30}$$

In this case, the tax on interest is unambiguously positive, while the tax on wealth may assume either sign and is negative for all sufficiently small consumption volatility. Further,

---

[17]A general expression for the optimal taxes is given in Appendix D.4.

[18]This is also the specification chosen by Jones and Kim [2018], who further assume logarithmic utility.

expressions (29) and (30) illustrate that in this special case, the optimal taxes on interest and wealth tend to zero smoothly as the degree of idiosyncratic risk vanishes.

Note that the expressions for taxes in equations (28) and (29) are highly sensitive to changes in the parameter $\underline{l}$, and in fact diverge as this quantity and the fraction of entrepreneurs vanish. Further, the parameter $\underline{l}$ is difficult to discipline empirically as it corresponds to the lowest amount of leisure physically possible for the entrepreneur. It is for this reason that I have emphasized the novel qualitative features of the optimal tax policy and the conceptually distinct roles played by each tax instead of emphasizing a particular calibration. Corollary 4.6 shows that the potentially large statutory rates do not necessarily translate into high revenue, which depends only on the parameter $\overline{\gamma}$ and the efficient volatility of consumption. Figure 5 depicts the revenue raised per unit of wealth for various values of the degree of risk aversion. On the left-hand side volatility is fixed at $\sigma = 0.2$ and leisure varies, while on the right-hand side leisure is fixed at the lowest value and the volatility varies. As expected, the revenue raised is increasing in volatility σ and decreasing in leisure. For the parameters considered in Section 2, the revenue raised varies betweenapproximately 1 and 2.5 percent, while in the logarithmic case, the revenue raised from the interest tax is completely offset by a tax/subsidy on wealth and these two taxes raise no revenue in the aggregate.[19]
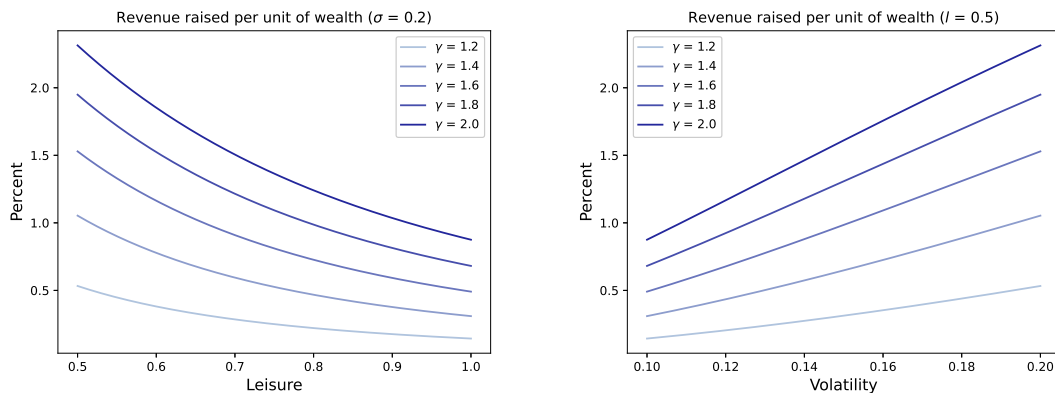

Figure 5: Revenue raised per unit of wealth.

Having derived expressions for the optimal linear taxes and computed some examples,

---

[19]Appendix E shows that the expression for revenue in Corollary 4.6 also holds in an implementation of the restricted-action allocations with competitive share markets wherever the latter is well-defined.

it is instructive to briefly return to a discussion of how these results differ from those in the literature. As mentioned above, the fact that the optimal linear taxation scheme calls for double taxation of capital income, at both the firm and individual level, is reminiscent of the seminal contribution of Albanesi [2006], who arrives at such a prescription within a two-period model with an exogenous distribution of physical capital. Since production occurs only once, the question of how capital ought to be distributed in the future never arises, and if capital could be reallocated, the planner would wish to assign it all to a single entrepreneur to minimize the costs associated with private information. The primary difference between the current paper and a dynamic extension of Albanesi [2006] is therefore that business income in the current paper depends on the owner's effort. This assumption, adopted to capture the empirical evidence of Smith et al. [2019], implies that it is simply not possible to reallocate productive resources across entrepreneurs. In the efficient allocations in this paper, even arbitrarily successful and rich entrepreneurs continue to exert effort as their firm grows, as only they are capable of affecting productivity growth. The above taxes simply ensure that these incentives to retain ownership and exert effort coincide imply that consumption growth coincides with that given in Section 3.

5   Conclusion

This paper characterizes a class of constrained efficient allocations in an economy with moral hazard and stochastic growth in business income. I characterize constrained-efficient allocations in an economy with persistent effects of effort and then use the insights from the mechanism design approach to derive optimal linear taxes in signaling equilibria.

The principal findings from each section were as follows. First, allowing effort to affect output growth rather than levels 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 standard preferences, agents with high realizations of shocks become too expensive to motivate and so are eventually retired. In contrast, in the model of this paper agents become richer because they experience high productivity growth, and so the benefits of further effort rise along with the costs. In Lemma 2.4 I provide sufficient conditions for this second force to overwhelm the first within a restricted class of contracts and provide numerical evidence that this often holds in the unrestricted optimal contract. In Proposition 3.5 I then illustrate the importance of this force by characterizing the associated upper tails of the stationary distributions of consumption and firm size (in

terms of output or employment), showing that they do in general differ and that the latter will typically be thicker than the former.

Second, I derived the optimal linear taxes on capital income and wealth in an environment in which agents may trade shares in their firms or save in a risk-free bond. In this case Proposition 4.5 uncovers a novel role for taxes when productivity depends on unobserved effort: a tax on interest income alters the incentives of owners to retain ownership of their firm, and hence to exert continued effort to improve productivity. The optimal linear taxation policy in this environment therefore calls for taxes on profits, risk-free interest, and wealth, serving three qualitatively distinct purposes. The tax on profits plays a redistributive role, the tax on risk-free savings affects the incentives for retaining equity and continued effort, and the wealth tax serves to implement the efficient level of consumption smoothing.

References

Hengjie Ai, Dana Kiku, and Rui Li. A Quantitative Model of Dynamic Moral Hazard. *The Review of Financial Studies*, page hhac059, August 2022. doi:10.1093/rfs/hhac059.

Stefania Albanesi. Optimal Taxation of Entrepreneurial Capital with Private Information. Technical Report w12419, National Bureau of Economic Research, August 2006.

Laurence Ales and Christopher Sleet. Taxing Top CEO Incomes. *American Economic Review*, 106(11):3331–3366, November 2016. doi:10.1257/aer.20151093.

Laurence Ales, Antonio Andrés Bellofatto, and Jessie Jiaxu Wang. Taxing Atlas: Executive Compensation, Firm Size, and Their Impact on Optimal Top Income Tax Rates. *Review of Economic Dynamics*, 26:62–90, October 2017. doi:10.1016/j.red.2017.02.007.

George-Marios Angeletos. Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, 10(1):1–30, January 2007. doi:10.1016/j.red.2006.11.001.

Andrew Atkeson and Robert E. Lucas, Jr. On Efficient Distribution With Private Information. *The Review of Economic Studies*, 59(3):427–453, July 1992. doi:10.2307/2297858.

Michael Best and Henrik Kleven. Optimal Income Taxation with Career Effects of Work Effort. SSRN Scholarly Paper ID 2051214, Social Science Research Network, Rochester, NY, April 2012.

Marco Cagetti and Mariacristina De Nardi. Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy*, 114(5):835–870, October 2006. ISSN 0022-3808. doi:10.1086/508032. URL https://www.journals.uchicago.edu/doi/full/10.1086/508032. Publisher: The University of Chicago Press.

Christophe Chamley. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, 54(3):607–622, 1986. doi:10.2307/1911310.

V. V. Chari and Patrick J. Kehoe. Optimal Fiscal and Monetary Policy. In *Handbook of Macroeconomics*, volume 1, pages 1671–1745. Elsevier, January 1999. doi:10.1016/S1574-0048(99)10039-9.

Michael Cooper, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick. Business in the United States: Who Owns It, and How Much Tax Do They Pay? *Tax Policy and the Economy*, 30(1):91–128, January 2016. doi:10.1086/685594. Publisher: The University of Chicago Press.

Jakša Cvitanić, Xuhu Wan, and Jianfeng Zhang. Optimal Compensation with Hidden Action and Lump-Sum Payment in a Continuous-Time Model. *Applied Mathematics and Optimization*, 59(1):99–146, February 2009. doi:10.1007/s00245-008-9050-0.

Peter Diamond and Emmanuel Saez. The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives*, 25(4):165–190, December 2011. doi:10.1257/jep.25.4.165.

Emmanuel Farhi and Iván Werning. Inequality and Social Discounting. *Journal of Political Economy*, 115(3):365–402, June 2007. doi:10.1086/518741.

Emmanuel Farhi and Iván Werning. Insurance and Taxation over the Life Cycle. *The Review of Economic Studies*, 80(2):596–635, April 2013. doi:10.1093/restud/rds048.

Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal Indirect and Capital Taxation. *The Review of Economic Studies*, 70(3):569–587, 2003. doi:10.1111/1467-937X.00256.

Mikhail Golosov, Maxim Troshkin, and Aleh Tsyvinski. Redistribution and Social Insurance. *American Economic Review*, 106(2):359–386, February 2016. doi:10.1257/aer.20111550.

Fatih Guvenen, Gueorgui Kambourov, Burhanettin Kuruscu, Sergio Ocampo-Diaz, and Daphne Chen. Use It or Lose It: Efficiency Gains from Wealth Taxation. Technical Report w26284, National Bureau of Economic Research, September 2019.

Zhiguo He. Optimal Executive Compensation When Firm Size Follows Geometric Brownian Motion. *The Review of Financial Studies*, 22(2):859–892, February 2009. doi:10.1093/rfs/hhn037.

Charles I. Jones and Jihee Kim. A Schumpeterian Model of Top Income Inequality. *Journal of Political Economy*, 126(5):1785–1826, June 2018. doi:10.1086/699190.

Kenneth L. Judd. Redistributive Taxation in a Simple Perfect Foresight Model. *Journal of Public Economics*, 28(1):59–83, October 1985. doi:10.1016/0047-2727(85)90020-9.

Marek Kapička and Julian Neira. Optimal Taxation with Risky Human Capital. *American Economic Journal: Macroeconomics*, 11(4):271–309, October 2019. doi:10.1257/mac.20160365.

Harold J. Kushner and Paul G. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer Science & Business Media, 2001. ISBN 978-0-387-95139-3.

Robert E. Lucas. On the Size Distribution of Business Firms. *The Bell Journal of Economics*, 9(2):508–523, 1978. doi:10.2307/3003596.

Miltiadis Makris and Alessandro Pavan. Taxation under Learning-by-Doing. *Journal of Political Economy*, January 2021. doi:10.1086/713745.

J. A. Mirrlees. An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies*, 38(2):175–208, 1971. doi:10.2307/2296779.

Christopher Phelan and Robert M. Townsend. Computing Multi-Period, Information-Constrained Optima. *The Review of Economic Studies*, 58(5):853–881, October 1991. doi:10.2307/2297941.

Thomas Phelan and Keyvan Eslami. Applications of Markov chain approximation methods to optimal control problems in economics. *Journal of Economic Dynamics and Control*, 143:104437, October 2022. ISSN 0165-1889. doi:10.1016/j.jedc.2022.104437.

Tom Phelan. On the Optimality of Differential Asset Taxation. *Working Paper*, (19-17R), September 2022. doi:10.26509/frbc-wp-201917r.

Sherwin Rosen. Authority, Control, and the Distribution of Earnings. *The Bell Journal of Economics*, 13(2):311–323, 1982. doi:10.2307/3003456.

Yuliy Sannikov. A Continuous-Time Version of the Principal-Agent Problem. *The Review of Economic Studies*, 75(3):957–984, July 2008. doi:10.1111/j.1467-937X.2008.00486.x.

Florian Scheuer. Entrepreneurial Taxation with Endogenous Entry. *American Economic Journal: Economic Policy*, 6(2):126–163, May 2014. doi:10.1257/pol.6.2.126.

Florian Scheuer and Iván Werning. The Taxation of Superstars*. *The Quarterly Journal of Economics*, 132(1):211–270, February 2017. doi:10.1093/qje/qjw036.

Ali Shourideh. Optimal Taxation of Wealthy Individuals. 2013 Meeting Paper 261, Society for Economic Dynamics, 2013.

Matthew Smith, Danny Yagan, Owen Zidar, and Eric Zwick. Capitalists in the Twenty-First Century. *The Quarterly Journal of Economics*, 134(4):1675–1745, November 2019. doi:10.1093/qje/qjz020.

Stefanie Stantcheva. Optimal Taxation and Human Capital Policies over the Life Cycle. *Journal of Political Economy*, 125(6):1931–1990, August 2017. ISSN 0022-3808. doi:10.1086/694291.

List of Figures

A   Recursive analysis

In this section I outline the arguments leading to the Hamilton-Jacobi-Bellman equation. For clarity, define the underlying filtered probability space to be $(\Omega, (\mathcal{F}_t)_{t \geq 0}, P)$, where $\Omega = C([0, \infty))$, P is the Wiener measure and $\mathcal{F}$ is the σ-algebra generated by the evaluation maps. In this case the

assumption that $l = (l_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, P)$ is progressively measurable is equivalent to the existence of functions $(\tilde{l}_t)_{t \geq 0}$ such that $\tilde{l}_t : C([0,t]) \to \mathbb{R}$ for each $t \geq 0$ and $l_t = \tilde{l}_t((\omega(s))_{0 \leq s \leq t})$ almost surely, for all $t \geq 0$. These definitions amount to assuming that the choices of the principal and agent at any date are functions of the history of quantities observed up until that date.

A.1 Incentive compatibility

For each l define $Z_t^l(\omega) := \omega(t) - \sigma^{-1} \int_0^t \mu_\theta(\tilde{l}_s((\omega(s'))_{0 \leq s' \leq s}))ds$ for $\omega \in \Omega$, and note that for any l and l' we have $dZ_t^{l'} = dZ_t^l - \sigma^{-1}[\mu_\theta(l'_t) - \mu_\theta(l_t)]$. We define P^l to be the measure under which Z^l is a Brownian motion on $(\Omega, \mathcal{F}, P^l)$ (see e.g. page 106 of Cvitanić et al. [2009]). Given the underlying Brownian motion $Z := (Z_t)_{t \geq 0}$, we define θ to be a (strong) solution to $d\theta_t = \sigma \theta_t dZ_t$ and note that

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

and so (θ, Z^l, P^l) is a (weak) solution to $dX_t = \mu_\theta(\tilde{l}(X.))X_t dt + \sigma X_t dW_t$. Finally, we write $\mathbb{E}^l$ for the associated expectation operator on the space of output paths.

Proof of Proposition 2.1. For any allocation (c, l^P) and agent strategy l, define $V^{c,l} := (V_t^{c,l})_{t \geq 0}$ by

$$V_t^{c,l} := \mathbb{E}^l \bigg[\int_0^\infty \rho e^{-\rho s} u(c_s, l_s) ds \bigg| \mathcal{F}_t \bigg]. \tag{31}$$

The law of iterated expectations ensures that $V_t^{c,l}$ is a martingale on $(\Omega, \mathcal{F}, P^l)$. The form of the martingale representation theorem given in Lemma 3.1 of Cvitanić et al. [2009] then implies the existence of S such that $V_t^{c,l} = \rho \sigma \int_0^t S_s dZ_s^l$ for all $t \geq 0$ almost surely, which gives the first claim. Now consider the case in which the agent adheres to an arbitrary l up until a fixed date t and adheres to l^P thereafter. In this case (31) becomes $V_t^{c,l} := \rho \int_0^t e^{-\rho s} u(c_s, l_s) ds + e^{-\rho t} U_t$, where U_t is utility if the agent adheres to l^P after t. Using $dZ_t^{l^P} = dZ_t^l - \sigma^{-1}[\mu_\theta(l_t^P) - \mu_\theta(l_t)]$ we have

$$
\begin{aligned}
dV_t^l &= \rho e^{-\rho t} u(c_t, l_t) dt + d\big(e^{-\rho t} U_t\big) \\
&= \rho e^{-\rho t}(u(c_t, l_t) dt - U_t) dt + e^{-\rho t}\Big[\rho(U_t - u(c_t, l_t^P))dt + \rho \sigma S_t dZ_t^{l^P}\Big] \\
&= \rho e^{-\rho t}\big(u(c_t, l_t) - u(c_t, l_t^P) + S_t[\mu_\theta(l_t) - \mu_\theta(l_t^P)]dt\big) + \rho \sigma e^{-\rho t} S_t dZ_t^l.
\end{aligned}
$$

Since $\mathbb{E}^l \Big[\int_0^t e^{-\rho s} S_s dZ_s^l \Big] = 0$ for all $t \geq 0$, we have

$$\mathbb{E}^l[V_t^l] = \hat{V}_0^l + \rho \mathbb{E}^l \bigg[\int_0^t e^{-\rho s}\big(S_s \mu_\theta(l_s) + u(c_s, l_s) - [S_s \mu_\theta(l_s^P) + u(c_s, l_s^P)]\big)ds \bigg]. \tag{32}$$

Since the expected utility of the agent is $\mathbb{E}^l[\lim_{t \to \infty} V_t^l]$, a recommendation l is incentive compatible if and only if it maximizes the integrand in (32) almost surely for all $t \geq 0$, which gives the result. $\square$

Remark A.1. *The strengthening of the martingale representation appearing in Cvitanić et al. [2009] and invoked above appears necessary because $\mathcal{F}$ is the filtration generated by the evaluation maps and is not necessarily equal to the natural filtration associated with Z^l.*

A.2 Laws of motion

We first assume $\gamma > 1$. In this case utility in consumption units and normalized utility are

$$z_t := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}} \qquad\qquad u_t := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}\theta_t^{-1}. \tag{33}$$

Recall $\overline{\gamma} := 1 - (1-\gamma)(1-\alpha)$ and $E(l) := \rho\alpha 1_{l<1}/[(1-\alpha)(\overline{\mu}_0 - \overline{\mu}_1)l]$ for any $l \in [\underline{l}, 1]$, and

$$u(c,l) = \frac{\left(c^{1-\alpha}l^\alpha\right)^{1-\gamma}}{1-\gamma} \qquad\qquad u_2(c,l) = \alpha l^{-1}\left(c^{1-\alpha}l^\alpha\right)^{1-\gamma}$$

$$\mu_\theta(l) = \overline{\mu}_0 - (\overline{\mu}_0 - \overline{\mu}_1)l \qquad\qquad \mu_\theta'(l) = -(\overline{\mu}_0 - \overline{\mu}_1)$$

if $l < 1$, and $\mu_\theta(1) = 0$. Proposition 2.1 shows that we may suppose

$$dU_t = \rho(U_t - u(c_t, l_t))dt + \sigma(1-\alpha)E(l_t)\left(c_t^{1-\alpha}l_t^\alpha\right)^{1-\gamma}dZ_t. \tag{34}$$

We write $c_t = \overline{c}_t[(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}$, for $\overline{c}_t$ interpreted as consumption per utility in consumption units, and note that $(c_t^{1-\alpha}l_t^\alpha)^{1-\gamma} = (\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}(1-\gamma)U_t$. The law of motion (34) becomes $dU_t/U_t = \mu_U dt + \sigma_U dZ_t$ where

$$\mu_U = \rho\left(1 - (\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}\right) \qquad\qquad \sigma_U = \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}(1-\overline{\gamma}). \tag{35}$$

We now wish to derive the laws of motion of z_t and u_t in (33). If $f(U_t) := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}$ then

$$f'(U_t) := \frac{1}{(1-\alpha)}[(1-\gamma)U_t]^{\frac{\overline{\gamma}}{1-\overline{\gamma}}} = \frac{U_t^{-1}}{(1-\overline{\gamma})}f(U_t)$$

$$f''(U_t) := \frac{\overline{\gamma}}{(1-\alpha)^2}[(1-\gamma)U_t]^{\frac{\overline{\gamma}}{1-\overline{\gamma}}-1} = \frac{\overline{\gamma}U_t^{-2}}{(1-\overline{\gamma})^2}f(U_t)$$

and so Ito's lemma and (35) imply $dz_t = \mu_z z_t dt + \sigma_z z_t dZ_t$, where

$$\mu_z = \frac{\mu_U}{1-\overline{\gamma}} + \frac{\sigma_U^2}{2}\frac{\overline{\gamma}}{(1-\overline{\gamma})^2} \qquad\qquad \sigma_z = \frac{\sigma_U}{1-\overline{\gamma}}. \tag{36}$$

Substituting (35) into (36) gives

$$\mu_z = \rho\left(\frac{1 - (\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right) + \frac{\overline{\gamma}\sigma^2}{2}E(l_t)^2(\overline{c}_t^{1-\alpha}l_t^\alpha)^{2-2\gamma}$$

$$\sigma_z = \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}. \tag{37}$$

Since $u_t = z_t/\theta_t$, using Ito's lemma for quotients[20] and (37) we have $du_t/u_t = \mu_u dt + \sigma_u dZ_t$ where

$$\mu_u = \rho\left(\frac{1 - (\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right) + \frac{\overline{\gamma}\sigma^2}{2}E(l_t)^2(\overline{c}_t^{1-\alpha}l_t^\alpha)^{2-2\gamma} - \mu_\theta(l_t)$$

$$+ \sigma_\theta(l_t)^2 - \sigma\sigma_\theta(l_t)E(l_t)(\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma}$$

$$\sigma_u = \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^\alpha)^{1-\gamma} - \sigma_\theta(l_t).$$

---

[20]If $dX_t/X_t = \mu_X dt + \sigma_X dZ_t$ and $dY_t/Y_t = \mu_Y dt + \sigma_Y dZ_t$ then $d(X_t/Y_t)/(X_t/Y_t) = (\mu_X - \mu_Y + \sigma_Y^2 - \sigma_X\sigma_Y)dt + (\sigma_X - \sigma_Y)dZ_t$. In the above we use this with $X_t := z_t$ and $Y_t := \theta_t$.

Factorization and simplification implies that

$$\mu_u = \rho\left(\frac{1 - (\bar{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}}{1 - \overline{\gamma}}\right) + (\overline{\gamma} - 1)\frac{\sigma^2}{2}E(l_t)^2(\bar{c}_t^{1-\alpha}l_t^{\alpha})^{2-2\gamma}$$
$$+ \frac{\sigma_\theta^2(l_t)}{2}(E(l_t)(\bar{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma} - 1)^2 - \mu_\theta(l_t) + \frac{\sigma_\theta^2(l_t)}{2} \tag{38}$$

which in turn implies (10). The above expressions fail to be well-defined when $\gamma = 1$, and so we treat this case separately. We define flow utility to be $u(c, l) = \alpha \ln c + (1 - \alpha)\ln l$, and so the analogue of (33) is $z_t = \exp(U/(1 - \alpha))$ and $u_t = \exp(U/(1 - \alpha))\theta^{-1}$. Writing $c_t = \bar{c}_t z_t$, we have $U_t - (1 - \alpha)\ln c_t = -(1 - \alpha)\ln\bar{c}_t$, and so (34) becomes

$$dU_t = \rho(-(1 - \alpha)\ln\bar{c}_t - \alpha\ln l_t)dt + \sigma(1 - \alpha)E(l_t)dZ_t =: \mu_U dt + \sigma_U dZ_t.$$

If $f(U_t) := \exp(U/(1 - \alpha))$, then $f'(U_t) = (1 - \alpha)^{-1}f(U_t)$ and $f''(U_t) = (1 - \alpha)^{-2}f(U_t)$, and so from Ito's lemma, we have $dz_t/z_t = \mu_z dt + \sigma_z dZ_t$, where

$$\mu_z = \frac{\mu_U}{1 - \alpha} + \frac{\sigma_U^2}{2(1 - \alpha)^2} = -\rho\left(\ln\bar{c}_t + \frac{\alpha}{1 - \alpha}\ln l_t\right) + \frac{\sigma^2}{2}E(l_t)^2$$
$$\sigma_z = \frac{\sigma_U}{1 - \alpha} = \sigma E(l_t). \tag{39}$$

Using Ito's lemma for quotients gives $du_t/u_t = \mu_u dt + \sigma_u dZ_t$, where $\sigma_u = \sigma(E(l_t) - 1_{l_t<1})$ and

$$\mu_u = -\rho\ln\bar{c}_t - \frac{\rho\alpha}{1 - \alpha}\ln l_t + \frac{\sigma^2}{2}E(l_t)^2 - \mu_\theta(l_t) + \sigma_\theta(l_t)^2 - \sigma^2 E(l_t)$$
$$= -\rho\ln\bar{c}_t - \frac{\rho\alpha}{1 - \alpha}\ln l_t - \mu_\theta(l_t) + \frac{\sigma_\theta(l_t)^2}{2}(E(l_t) - 1)^2 + \frac{\sigma_\theta(l_t)^2}{2}. \tag{40}$$

For the purpose of characterizing the stationary distribution, it is instructive to also write out the laws of motion for the logarithmic transformations of the state variables u and θ. Defining $y_t := f(\theta_t) := \ln\theta_t$, and using $f'(\theta) = 1/\theta$ and $f''(\theta) = -1/\theta^2$, Ito's Lemma implies that

$$df(\theta_t) = \left(\mu_\theta\theta_t f'(\theta_t) + \sigma_\theta^2\theta_t^2 f''(\theta_t)/2\right)dt + \sigma_\theta\theta_t f'(\theta_t)dZ_t.$$

It follows that y_t and the similarly transformed $w_t := \ln(u_t)$ satisfy

$$dy_t = \mu_y dt + \sigma_y dZ_t$$
$$(\mu_y, \sigma_y) = (\mu_\theta - \sigma_\theta^2/2, \sigma_\theta)$$
$$dw_t = \mu_w dt + \sigma_w dZ_t \tag{41}$$
$$(\mu_w, \sigma_w) = (\mu_u - \sigma_u^2/2, \sigma_u).$$

I will use the logarithmically transformed system (w, y) when computing stationary distributions because this aids with accuracy when both are (approximately) Pareto distributed. However, because the value function is approximately linear (the restricted-action value function is exactly linear) I use the original variable u when computing the value function.

A.3 Hamilton-Jacobi-Bellman equation

Using the original choice variable (c, l), the law of motion of utility (34), and the law of motion of productivity (3), the Hamilton-Jacobi-Bellman equation is given by

$$\rho V(U, \theta) = \max_{\substack{c \geq 0 \\ l \in [\underline{l}, 1]}} \theta 1_{l < 1} - c + \rho(U - u(c, l)) \frac{\partial V}{\partial U} + \frac{\sigma^2}{2} \left((1 - \alpha) E(l) \left(c^{1-\alpha} l^\alpha \right)^{1-\gamma} \right)^2 \frac{\partial^2 V}{\partial U^2}$$
$$+ \mu_\theta(l) \theta \frac{\partial V}{\partial \theta} + \frac{\sigma_\theta(l)^2}{2} \theta^2 \frac{\partial^2 V}{\partial \theta^2} + \sigma(1 - \alpha) E(l) \left(c^{1-\alpha} l^\alpha \right)^{1-\gamma} \sigma_\theta(l) \theta \frac{\partial^2 V}{\partial U \partial \theta}. \tag{42}$$

If $\gamma > 1$, then in terms of the variables $(\bar{c}, l)$, where $c = \bar{c}[(1 - \gamma) U_t]^{\frac{1}{1-\bar{\gamma}}}$, (42) becomes

$$\rho V(U, \theta) = \max_{\substack{\bar{c} \geq 0 \\ l \in [\underline{l}, 1]}} \theta 1_{l < 1} - \bar{c}[(1 - \gamma) U]^{\frac{1}{1-\bar{\gamma}}} + \rho \left(1 - \left(\bar{c}^{1-\alpha} l^\alpha \right)^{1-\gamma} \right) U \frac{\partial V}{\partial U}$$
$$+ \frac{\sigma^2}{2} \left((1 - \bar{\gamma}) E(l) \left(\bar{c}^{1-\alpha} l^\alpha \right)^{1-\gamma} \right)^2 U^2 \frac{\partial^2 V}{\partial U^2} + \mu_\theta(l) \theta \frac{\partial V}{\partial \theta} \tag{43}$$
$$+ \frac{\sigma_\theta(l)^2}{2} \theta^2 \frac{\partial^2 V}{\partial \theta^2} + \sigma(1 - \bar{\gamma}) E(l) \left(\bar{c}^{1-\alpha} l^\alpha \right)^{1-\gamma} U \sigma_\theta(l) \theta \frac{\partial^2 V}{\partial U \partial \theta}$$

where we used $(c^{1-\alpha} l^\alpha)^{1-\gamma} = (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}(1 - \gamma) U$. In the case of logarithmic utility, $(\gamma = 1)$, in terms of the variables $(\bar{c}, l)$, where $c = \bar{c} \exp(U/(1 - \alpha))$, (43) becomes

$$\rho V(U, \theta) = \max_{\substack{\bar{c} \geq 0 \\ l \in [\underline{l}, 1]}} \theta 1_{l < 1} - \bar{c} \exp(U/(1 - \alpha)) + \rho(-(1 - \alpha) \ln \bar{c} - \alpha \ln l) \frac{\partial V}{\partial U}$$
$$+ \frac{\sigma^2}{2} (1 - \alpha)^2 E(l)^2 \frac{\partial^2 V}{\partial U^2} + \mu_\theta(l) \theta \frac{\partial V}{\partial \theta} + \frac{\sigma_\theta(l)^2 \theta^2}{2} \frac{\partial^2 V}{\partial \theta^2} + \sigma(1 - \alpha) E(l) \sigma_\theta(l) \theta \frac{\partial^2 V}{\partial U \partial \theta}. \tag{44}$$

Proposition A.1. *When $\gamma > 1$ the solution to the Hamilton-Jacobi-Bellman equation (43) is of the form $V(U, \theta) = v(u)\theta$ for some function v solving*

$$\rho v(u) = \max_{\substack{\bar{c} \geq 0 \\ l \in [\underline{l}, 1]}} 1_{l < 1} - \bar{c}u + \rho \left(\frac{1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1 - \bar{\gamma}} \right) uv'(u) + \frac{\bar{\gamma} \sigma^2}{2} E(l)^2 \left(\bar{c}^{1-\alpha} l^\alpha \right)^{2-2\gamma} uv'(u)$$
$$+ \mu_\theta(l)(v(u) - uv'(u)) + \frac{1}{2} \left(\sigma E(l) \left(\bar{c}^{1-\alpha} l^\alpha \right)^{1-\gamma} - \sigma_\theta(l) \right)^2 u^2 v''(u).$$

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

$$u = [(1 - \gamma) U]^{\frac{1}{1-\bar{\gamma}}} \theta^{-1} \qquad\qquad U = \frac{(u\theta)^{1-\bar{\gamma}}}{1 - \gamma}$$

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

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

$$\frac{\partial V}{\partial U} = \frac{\partial u}{\partial U} v'(u)\theta = \frac{\theta^{\bar{\gamma}} u^{\bar{\gamma}}}{1 - \alpha} v'(u)$$
$$\frac{\partial V}{\partial \theta} = v(u) + \frac{\partial u}{\partial \theta} v'(u)\theta = v(u) - uv'(u)$$

and

$$\frac{\partial^2 V}{\partial \theta^2} = \frac{\partial}{\partial \theta}[v(u) - v'(u)u] = \frac{\partial u}{\partial \theta}[v'(u) - v'(u) - v''(u)u] = v''(u)\frac{u^2}{\theta}$$

$$\frac{\partial^2 V}{\partial U^2} = \frac{\theta^{\overline{\gamma}}}{1-\alpha}\frac{\partial}{\partial U}[u^{\overline{\gamma}}v'(u)] = \frac{\theta^{\overline{\gamma}}}{1-\alpha}\frac{\partial u}{\partial U}[\overline{\gamma}u^{\overline{\gamma}-1}v'(u) + u^{\overline{\gamma}}v''(u)] = \frac{(u\theta)^{2\overline{\gamma}-2}}{(1-\alpha)^2}[\overline{\gamma}uv'(u) + u^2v''(u)]\theta$$

$$\frac{\partial^2 V}{\partial U \partial \theta} = \frac{\partial}{\partial U}(v(u) - uv'(u)) = \frac{\partial u}{\partial U}[-uv''(u)] = -\frac{u^{1+\overline{\gamma}}v''(u)}{1-\alpha}\theta^{\overline{\gamma}-1}.$$

Substitution into (43) gives

$$\rho v(u)\theta = \max_{\substack{\overline{c} \geq 0 \\ l \in [\underline{l}, 1]}} (1_{l<1} - \overline{c}u)\theta + \rho\left(1 - (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}\right)U\frac{\theta^{\overline{\gamma}}u^{\overline{\gamma}}}{1-\alpha}v'(u)$$

$$+ \frac{\sigma^2}{2}E(l)^2(1-\overline{\gamma})^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma}U^2\frac{(u\theta)^{2\overline{\gamma}-2}}{(1-\alpha)^2}[\overline{\gamma}uv'(u) + u^2v''(u)]\theta$$

$$+ \mu_\theta(l)\theta(v(u) - uv'(u)) + \frac{\sigma_\theta(l)^2\theta}{2}v''(u)u^2 - \sigma\sigma_\theta(l)E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}v''(u)u^{1+\overline{\gamma}}(1-\gamma)U\theta^{\overline{\gamma}}.$$

Dividing by θ and using $u^{1-\overline{\gamma}} = (1-\gamma)U\theta^{\overline{\gamma}-1}$ and (7) then gives

$$\rho v(u) = \max_{\substack{\overline{c} \geq 0 \\ l \in [\underline{l}, 1]}} 1_{l<1} - \overline{c}u + \rho\left(\frac{1 - (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right)uv'(u) + \frac{\sigma^2}{2}E(l)^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma}[\overline{\gamma}uv'(u) + u^2v''(u)]$$

$$+ \mu_\theta(l)(v(u) - uv'(u)) + \frac{\sigma_\theta(l)^2}{2}u^2v''(u) - \sigma\sigma_\theta(l)E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}u^2v''(u)$$

which simplifies as claimed upon factorization. $\square$

We have the following analogue in the case of logarithmic utility. The proof is almost identical to that of Proposition A.1 and so omitted.

Proposition A.2. *When utility is logarithmic, the solution to the Hamilton-Jacobi-Bellman equation* (44) *is of the form* $V(U, \theta) = v(u)\theta$ *for some function* v *solving*

$$\rho v(u) = \max_{\substack{\overline{c} \geq 0 \\ l \in [\underline{l}, 1]}} 1_{l<1} - \overline{c}u + \rho\left(-\ln \overline{c} - \frac{\alpha \ln l}{1-\alpha}\right)uv'(u) + \frac{\sigma^2}{2}E(l)^2uv'(u)$$

$$+ \mu_\theta(l)(v(u) - uv'(u)) + \frac{\sigma^2}{2}(E(l) - 1_{l<1})^2u^2v''(u).$$

A.4 Restricted value and policy functions

Proof of Proposition 2.3. Using Proposition A.1, substituting the form $v \equiv \overline{v}_{\mathrm{r}}(l)u + \underline{v}_{\mathrm{r}}(l)$ and equating the linear and constant parts gives $\underline{v}_{\mathrm{r}}(l) = 1_{l<1}/(\rho - \mu_\theta(l))$ and $\overline{v}_{\mathrm{r}}(l)$ solves

$$\rho\overline{v}_{\mathrm{r}}(l) = \max_{\overline{c} \geq 0} -\overline{c} + \rho\left(\frac{1 - (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right)\overline{v}_{\mathrm{r}}(l) + \frac{\overline{\gamma}\sigma^2}{2}E(l)^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma}\overline{v}_{\mathrm{r}}(l). \tag{45}$$

Note that the right-hand side is concave in $\bar{c}$ and diverges to $-\infty$ for any $\bar{v}_{\mathrm{r}}(l) < 0$. Consequently, if a solution to (45) exists then the optimal consumption level satisfies a first-order condition and we therefore arrive at the pair of equations

$$\rho\bar{v}_{\mathrm{r}}(l) = -\bar{c} + \rho\left(\frac{1 - (\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1 - \overline{\gamma}}\right)\bar{v}_{\mathrm{r}}(l) + \frac{\overline{\gamma}\sigma^2}{2}E(l)^2\left(\bar{c}^{1-\alpha}l^\alpha\right)^{2-2\gamma}\bar{v}_{\mathrm{r}}(l)$$

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

Equating the two expressions for $\bar{c}$ and dividing by $\bar{v}_{\mathrm{r}}(l)$ gives a quadratic in $x := (\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}$

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

which simplifies to the claimed quadratic. Further, for the maintained assumption $\overline{\gamma} \geq 1$, this quadratic has only one positive solution (and hence only one sensible solution), which corresponds to the solution to (45). Using the fact that $c_t = \bar{c}u_t\theta_t = \bar{c}z_t$ for z_t as defined in Appendix A.2, the expression (37) then gives

$$\rho\bar{v}_{\mathrm{r}}(l) = -\bar{c} + \rho\left(\frac{1 - x}{1 - \overline{\gamma}}\right)\bar{v}_{\mathrm{r}}(l) + \frac{\overline{\gamma}\sigma^2}{2}E(l)^2 x^2\bar{v}_{\mathrm{r}}(l) = -\bar{c} + \mu_c(l)\bar{v}_{\mathrm{r}}(l).$$

which rearranges to give $\bar{v}_{\mathrm{r}}(l)$. Finally, note that $\bar{v}_{\mathrm{r}}(l)$ solves $\rho\bar{v} = T(\bar{v})$, where

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

Since $\overline{\gamma}, \gamma > 1$, T is increasing in l whenever $\bar{v}$ is negative since it is the pointwise maxima of increasing functions. The fixed-point will then be increasing in l provided the right-hand side is convex in $\bar{v}$, which is true as it is the pointwise maxima of affine functions. $\square$

The case of logarithmic utility corresponds to $\overline{\gamma} = 1$ and so the expressions in Propostion 2.3 are not well-defined. In this case we define $V(U, \theta) = v(u)\theta$ as before but now set $u = \exp(U/(1-\alpha))\theta^{-1}$.

Lemma A.3. *In the case of logarithmic utility the restricted-action value and policy functions are*

$$v_{\mathrm{r}}(u; l) = \frac{1_{l<1}}{\rho - \mu_\theta(l)} - \frac{1}{\rho}\exp\left(\sigma^2 E(l)^2/[2\rho]\right)l^{-\frac{\alpha}{1-\alpha}}u$$

$$c_{\mathrm{r}}(u; l) = \exp\left(\sigma^2 E(l)^2/[2\rho]\right)l^{-\frac{\alpha}{1-\alpha}}u. \tag{46}$$

Proof. Using Proposition A.2, substituting the form $v_{\mathrm{r}} \equiv \bar{v}_{\mathrm{r}}(l)u + \underline{v}_{\mathrm{r}}(l)$ and equating the linear and constant parts gives $\underline{v}_{\mathrm{r}}(l) = 1/(\rho - \mu_\theta(l))$ and $\bar{v}_{\mathrm{r}}(l)$ solves

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

The first-order condition gives $\bar{c} = \rho[-\bar{v}_{\mathrm{r}}(l)]$, and the above becomes $\ln(-\rho\bar{v}_{\mathrm{r}}(l)) = -\alpha(1-\alpha)^{-1}\ln l + \sigma^2 E(l)^2/[2\rho]$, which rearranges as claimed. $\square$

B Stationary distribution proofs

B.1 Planner preferences

Defining $\Omega' := \mathbb{R} \times \Theta \times \{E, W\}$, aggregate consumption, output and aggregate labor assigned to entrepreneurs at any date $t \geq 0$ are then,

$$C_t := e^{-\rho_D t}\underline{C}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} C_t^T dT$$

$$\underline{C}_t := \int_{\Omega'} \mathbb{E}[c_t^{v,\theta,i}]\Phi(d\omega), \qquad C_t^T := \sum_{i=E,W} \eta_i \mathbb{E}[c_t^{T,i}]$$

$$Y_t := e^{-\rho_D t}\underline{Y}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} Y_t^T dT$$

$$\underline{Y}_t := \int_{\Omega'} \mathbb{E}[F(\theta_t^{v,\theta,i}, L_t^{v,\theta,i})]\Phi(d\omega), \qquad Y_t^T := \eta_E \mathbb{E}[F(\theta_t^{T,E}, L_t^{T,E})]$$

$$L_t := e^{-\rho_D t}\underline{L}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} L_t^T dT$$

$$\underline{L}_t := \int_{\Omega} \mathbb{E}[L_t^{v,\theta,E}]\Phi(d\omega), \qquad L_t^T = \eta_E \mathbb{E}[L_t^{T,E}].$$

Since the planner only weights workers, the flow utility at date t is

$$U_t := e^{-\rho_D t}\underline{U}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} U_t^T dT$$

$$\underline{U}_t := \int_{\Omega'} \mathbb{E}[u(c_t^{v,\theta,i}, l_t^{v,\theta,i})]\Phi(d\omega), \qquad U_t^T := (1 - \eta_E)\mathbb{E}[u(c_t^{T,W}, l_t^{T,W})]$$

Lemma B.1. *The preferences of the planner are represented by the function $\int_0^\infty e^{-\rho_S t} U_t dt$.*

Proof. Proceeding from first principles, the objective of the planner is

$$U^P = \int_{\Omega'} \mathbb{E}\left[\int_0^\infty e^{-\rho t} u(c_t^{v,\theta,i}, l_t^{v,\theta,i})dt\right]\Phi(d\omega)$$

$$+ \rho_D \int_0^\infty e^{-\rho_S T}(1 - \eta_E)\mathbb{E}\left[\int_T^\infty e^{-\rho(t-T)} u(c_t^{T,W}, l_t^{T,W})dt\right]dT$$

$$= \int_0^\infty e^{-\rho_S t} \int_{\Omega'} e^{-\rho_D t}\mathbb{E}[u(c_t^{v,\theta,i}, l_t^{v,\theta,i})]dt\Phi(d\omega)dt$$

$$+ \rho_D(1 - \eta_E)\int_0^\infty e^{-\rho_S t} \int_0^t e^{-\rho_D(t-T)}\mathbb{E}[u(c_t^{T,W}, l_t^{T,W})]dTdt$$

where I used $e^{-\rho_S T}e^{-\rho(t-T)} = e^{-\rho_S t}e^{-\rho_D(t-T)}$, which gives the result upon simplification. □

B.2 Reduction to principal-agent problem

The Lagrangian associated with the relaxed planner's problem is

$$\mathcal{L} = \int_0^\infty e^{-\rho_S t}\big(U_t + \lambda_R[Y_t - C_t + \lambda_L(\overline{L} - L_t)]\big)dt.$$

Using the above expressions we may then expand this as

$$\mathcal{L} = \lambda_R\lambda_L\overline{L} + \int_0^\infty e^{-\rho t}(\underline{U}_t + \lambda_R[\underline{Y}_t - \lambda_L\underline{L}_t - \underline{C}_t])dt$$
$$+ \rho_D\int_0^\infty e^{-\rho_S T}\int_T^\infty e^{-\rho(t-T)}\big(U_t^T + \lambda_R[Y_t^T - \lambda_L L_t^T - C_t^T]\big)dtdT$$

where I again used $e^{-\rho_S T}e^{-\rho(t-T)} = e^{-\rho_S t}e^{-\rho_D(t-T)}$ and interchanged the order of integration. For any $T \geq 0$ I write $\mathcal{A}_T$ for the subset of allocations associated with agents born at time T, and $\mathcal{A}_T^{IC}$ for the associated incentive compatible allocations. For a given multiplier the task of the planner choosing quantities for agents born at date T is

$$\max_{\mathcal{A}_T^{IC}} \int_0^\infty e^{-\rho t}(U_t^T + \lambda_R[Y_t^T - \lambda_L L_t^T - C_t^T])dt.$$

Since the planner weights only workers, he provides all agents with a common utility level. For entrepreneurs the problem becomes $\max_U\ U + \lambda_R V(U, \overline{Z}(\lambda_L))$, where V is the value function of the principal in Section 2. The problem of the planner facing a given generation is therefore

$$V_\lambda^G = \max_{\substack{U_E, U_W < 0 \\ U_E \geq U_W}} \eta_E\big(U_E + \lambda_R\overline{Z}(\lambda_L)V(U_E\overline{Z}(\lambda_L)^{\overline{\gamma}-1}, 1)\big) + (1 - \eta_E)\Big(U_W - \lambda_R[(1-\gamma)U_W]^{\frac{1}{1-\overline{\gamma}}}\Big)$$
$$= \max_{y \geq 0} \overline{Z}(\lambda_L)^{1-\overline{\gamma}}\frac{y^{1-\overline{\gamma}}}{1-\gamma} + \lambda_R\overline{Z}(\lambda_L)[\eta_E v(y) - (1-\eta_E)y].$$

The task of the planner is to therefore pick a value of $y := [(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}/\overline{Z}(\lambda_L)$ and to then follow the recommendations of the principal with normalized utility given by y.

B.3 Kolmogorov forward equation for joint law

The Kolmogorov forward equation gives the evolution of the joint density of utility and productivity. The following shows that the summary measure given in Definition 3.5 solves a version of the Kolmogorov forward equation for a single variable. For a proof of the following, see Appendix B.3.

Lemma B.2. *If $(u_t, \theta_t)_{t \geq 0}$ evolves according to $(du_t, d\theta_t) = (\mu_u u_t dt + \sigma_u u_t dZ_t, \mu_\theta \theta_t dt + \sigma_\theta \theta_t dZ_t)$ for some $\mu_u, \sigma_u, \mu_\theta$ and σ_θ, then m solves the ordinary differential equation*

$$0 = -(\rho_D - \mu_\theta(u))m(u) - [(\mu_u(u) + \sigma_\theta(u)\sigma_u(u))um(u)]' + \frac{1}{2}[\sigma_u^2(u)u^2 m(u)]''. \tag{47}$$

The homogeneity of the policy functions and the exponential growth of productivity ensure that we need only solve for the density of a single variable, referred to as the summary measure.[21]

Proof of Lemma B.2. The process $(\theta_t, u_t)_{t \geq 0}$ is a diffusion process driven by the same Brownian motion and so away from (θ_0, u_0) the joint density satisfies the Kolmogorov forward equation

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

First note that the stationary distribution $\overline{\Phi}$ satisfies $\lim_{\theta \to \infty} \theta \overline{\Phi}(\theta, u) = 0$ for all u, since we have assumed that $\rho_D > \mu_\theta(l)$ for all $l \in [\underline{l}, 1]$. For any smooth f vanishing at zero and satisfying $\lim_{\theta \to \infty} \theta f(\theta) = 0$, 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

$$\mu_\theta(u) \int_0^\infty \theta \frac{\partial}{\partial \theta}[\theta \Phi(\theta, u, t)] d\theta = -\mu_\theta(u) m(u, t)$$

$$-\int_0^\infty \theta \frac{\partial}{\partial u}[\mu_u(u) \Phi(\theta, u, t)] d\theta = -\frac{\partial}{\partial u}[\mu_u(u) m(u, t)]$$

$$\frac{\sigma_\theta^2(u)}{2} \int_0^\infty \theta \frac{\partial^2}{\partial \theta^2}[\theta^2 \Phi(\theta, u, t)] d\theta = 0$$

$$\int_0^\infty \theta \frac{\partial^2}{\partial \theta \partial u}[\theta \sigma_u(u) \Phi(\theta, u, t)] d\theta = -\frac{\partial}{\partial u}[\sigma_u(u) m(u, t)]$$

$$\int_0^\infty \theta \frac{\partial^2}{\partial u^2}[\sigma_u^2(u) \Phi(\theta, u, t)] d\theta = \frac{\partial^2}{\partial u^2}[\sigma_u^2(u) m(u, t)].$$

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

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

which is equivalent to

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

Setting the partial derivative with respect to time to zero then gives (47). $\square$

---

[21] I owe the following observation to a discussion with Hengjie Ai.

B.4 Restricted-action distributions

For convenience the following recalls a well-known fact regarding killed diffusion processes.

Lemma B.3. *The stationary distribution of a stochastic process that dies at rate δ_X, is injected at some $\overline{X} > 0$, and otherwise evolves according to $dX_t = \mu_X X_t dt + \sigma_X X_t dZ_t$, is given by*

$$f(x) = Ax^{\alpha_X^+ - 1}1_{x \leq \overline{X}} + Bx^{\alpha_X^- - 1}1_{x > \overline{X}}$$

where $\alpha_X^\pm = \overline{\mu}_X/\sigma_X^2 \pm \sqrt{(\overline{\mu}_X/\sigma_X^2)^2 + 2\delta_X/\sigma_X^2}$, the constants A and B are chosen such that the density is continuous and integrates to unity, and $\overline{\mu}_X = \mu_X - \sigma_X^2/2$ for brevity.

Proof of Proposition 3.5. The proof proceeds by combining Lemma B.3 with the policy functions for consumption given in Proposition 2.3. From (37) we have $dc_t = \mu_c c_t dt + \sigma_c c_t dZ_t$, where

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

where $(\overline{\gamma} - 1)(\overline{\gamma} - 1/2)\sigma^2 E(l)^2 x^2 = \rho(1-x)$. Using the defining equation for $x(l)$ we then have

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

Consequently, $\mu_c - \sigma_c^2/2 = -\overline{\gamma}\sigma^2 E(l)^2 x^2/2$. Using Lemma B.3 the tails are then

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

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

It remains to determine the constants A and B appearing in the density in the statement of Lemma B.3. In order for the density to be continuous, the constants A and B must satisfy $A\overline{X}^{\alpha_X^+ - 1} = B\overline{X}^{\alpha_X^- - 1}$ and hence

$$A = B\overline{X}^{\alpha_X^- - \alpha_X^+}. \tag{48}$$

Now, we need B to be such that the density integrate to unity. This requires

$$1 = A\int_0^{\overline{X}} x^{\alpha_X^+ - 1}dx + B\int_{\overline{X}}^\infty x^{\alpha_X^- - 1}dx = \frac{A}{\alpha_X^+}\overline{X}^{\alpha_X^+} + \frac{B}{\alpha_X^-}\overline{X}^{\alpha_X^-}.$$

Using equation (48), we then require

$$1/B = \frac{\overline{X}^{\alpha_X^- - \alpha_X^+}}{\alpha_X^+}\overline{X}^{\alpha_X^+} + \frac{1}{\alpha_X^-}\overline{X}^{\alpha_X^-} = (1/\alpha_X^+ - 1/\alpha_X^-)\overline{X}^{\alpha_X^-}$$

and hence

$$B = \frac{\overline{X}^{-\alpha_X^-}}{(1/\alpha_X^+ - 1/\alpha_X^-)}. \tag{49}$$

The cdf for the above process is then given by $F(x) = \int_0^X f(x)dx$. For $X \leq \overline{X}$ we have $F(X) = AX^{\alpha_X^+}/\alpha_X^+$, and in generalequals

$$F(X) = \int_0^X \left(Ax^{\alpha_X^+ - 1} 1_{x \leq \overline{X}} + Bx^{\alpha_X^- - 1} 1_{x > \overline{X}} \right) dx$$

$$= A \min(X, \overline{X})^{\alpha_X^+}/\alpha_X^+ + B \min(X, \overline{X}) \left[X^{\alpha_X^-}/\alpha_X^- - \overline{X}^{\alpha_X^-}/\alpha_X^- \right].$$

Useing equations (48) and (49), this simplifies to

$$F(X) = B\overline{X}^{\alpha_X^- - \alpha_X^+} \min(X, \overline{X})^{\alpha_X^+}/\alpha_X^+ + B1_{X \geq \overline{X}} \left[X^{\alpha_X^-}/\alpha_X^- - \overline{X}^{\alpha_X^-}/\alpha_X^- \right]$$

$$= B \left(\min(1, X/\overline{X})/\alpha_X^+ + 1_{X \geq \overline{X}} \left[(X/\overline{X})^{\alpha_X^-} - 1 \right]/\alpha_X^- \right) \overline{X}^{\alpha_X^-}$$

$$= \frac{\alpha_X^- \alpha_X^+}{\alpha_X^- - \alpha_X^+} \left((X/\overline{X})^{\alpha_X^+}/\alpha_X^+ + 1_{X \geq \overline{X}} \left[(X/\overline{X})^{\alpha_X^-} - 1 \right]/\alpha_X^- \right).$$

The above expressions will serve as a check on the accuracy of the numerical method used to compute the stationary distribution.

C   Additional figures

In this appendix I compute several figures that complement those given in the main text.

C.1   Principal-agent problem

Figure 6 computes analogues of Figure 1 and Figure 2 when utility is logarithmic, which corresponds to the limit $\gamma \to 1$.
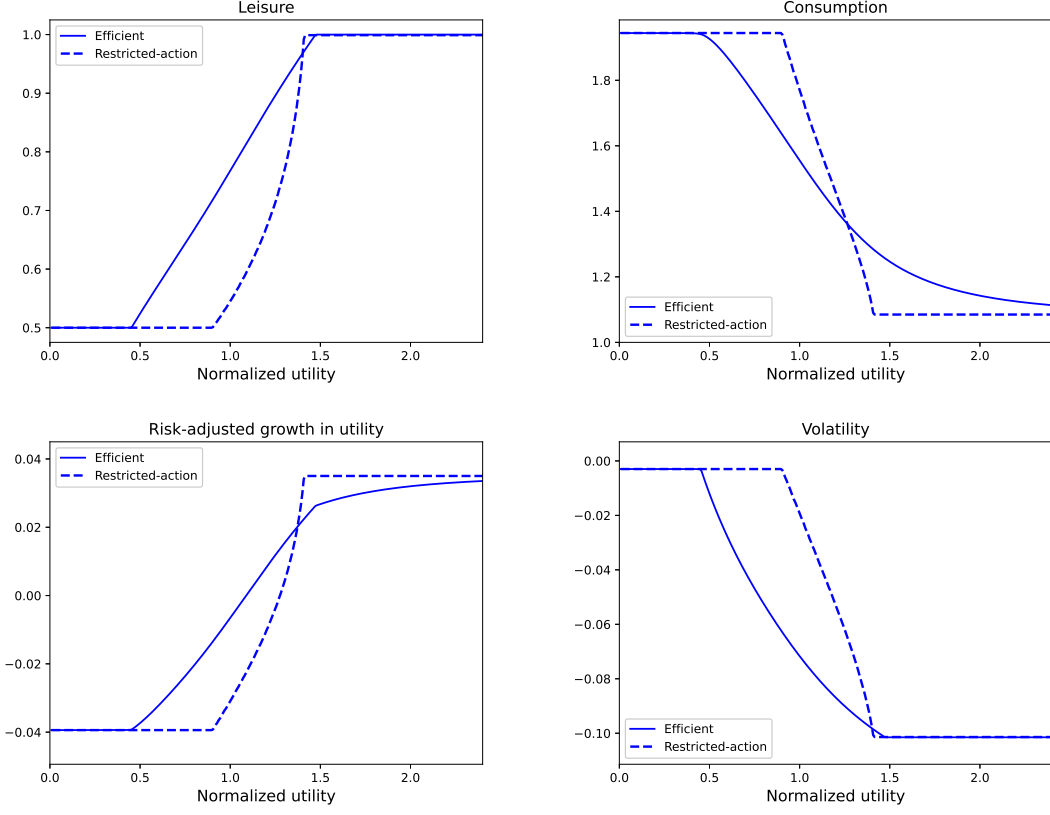
Figure 6: Policy functions and dynamics of normalized utility for logarithmic utility.

D   Optimal linear taxes

D.1   Discrete-time analogue

The continuous-time agent problem is the limit of discrete-time environments of the following form:

- At t the agent has wealth a_t. She places b_t in a risk-free bond earning after-tax return $(1 - \tau_s)(r + \rho_D)$ and uses the remaining $\iota_t := a_t - b_t$ to purchase shares at price p_t.

- At $t+\Delta$ shareholders receive after-tax dividends $\Delta(1-\tau_d)\overline{Z}\theta_t$ from the firm's output produced over $[t, t + \Delta]$. By $t + \Delta$ productivity has grown to $\overline{Z}\theta_{t+\Delta}$ and the price is $p_{t+\Delta}$.

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

$$
\begin{aligned}
a_{t+\Delta} &= a_t + \text{interest} - \text{consumption} + \text{dividends} + \text{capital gains} - \text{wealth tax} \\
&= a_t + (1 - \tau_s)(r + \rho_D)\Delta b_t - \Delta c_t + \Delta(1 - \tau_d)\overline{Z}\theta_t x_t + (1 - \tau_{cg})(p_{t+\Delta} - p_t)x_t - \tau_a \Delta a_t \\
&= a_t + \Delta\big[[(1 - \tau_s)(r + \rho_D) - \tau_a]a_t - c_t + [(1 - \tau_d)\overline{Z}\theta_t/p_t - (1 - \tau_s)(r + \rho_D)]\iota_t\big] \\
&\quad + (p_{t+\Delta}/p_t - 1)(1 - \tau_{cg})\iota_t.
\end{aligned}
\tag{50}
$$

The approximate law of motion for the price is $p_{t+\Delta}/p_t - 1 = \mu_\theta(l)\Delta + \sigma\sqrt{\Delta}dX$ where dX is mean zero i.i.d. across time and assumes values ± 1. Substituting into (50) and taking limits as $\Delta \to 0$

$$
\begin{aligned}
da_t &= \big[[(1 - \tau_s)(r + \rho_D) - \tau_a]a_t - c_t + \big((1 - \tau_d)\overline{Z}\theta_t/p_t + (1 - \tau_{cg})\mu_\theta(l_t) - (1 - \tau_s)(r + \rho_D)\big)\iota_t\big]dt \\
&\quad + \sigma(1 - \tau_{cg})\iota_t dZ_t.
\end{aligned}
$$

If $\hat{l}(u) \equiv \hat{l}$ then $p_t = (1 - \tau_d)\overline{Z}\theta_t/[r + \rho_D - \mu_\theta(\hat{l})]$, and so

$$
da_t = \Big[[(1 - \tau_s)(r + \rho_D) - \tau_a]a_t - c_t + \Big(\tau_s(r + \rho_D) + (1 - \tau_{cg})\mu_\theta(l_t) - \mu_\theta(\hat{l})\Big)\iota_t\Big]dt + \sigma(1 - \tau_{cg})\iota_t dZ_t
$$

which is the expression implied by (16) and (17).

D.2 Individual decision problems

Proof of Lemma 4.1. Given $(\bar{\iota}, \hat{l})$, I guess-and-verify that the value function of the agent is of the form $V(a) = \overline{V}a^{1-\overline{\gamma}}/(1 - \gamma)$ for some $\overline{V}$. Substituting this form into the Hamilton-Jacobi-Bellman equation associated with the problem in Definition 4.1 gives

$$
\begin{aligned}
\frac{\rho\overline{V}}{1 - \gamma} = \max_{\substack{\bar{c} \geq 0 \\ l \in [\underline{l}, 1]}} &\frac{\rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1 - \gamma} + [-\tau_a - \bar{c} + (1 - \tau_s)\rho](1 - \alpha)\overline{V} \\
&+ \Big(\big[\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s\rho\big]\bar{\iota} - \overline{\gamma}\sigma^2\bar{\iota}^2/2\Big)(1 - \alpha)\overline{V}.
\end{aligned}
\tag{51}
$$

The right-hand side of the Bellman equation (51) is concave, and so the maximum may be found by solving the first-order conditions for consumption and leisure,

$$
\begin{aligned}
\rho\bar{c}^{(1-\alpha)(1-\gamma)-1}l^{\alpha(1-\gamma)} &= \overline{V} \\
\rho\bar{c}^{(1-\alpha)(1-\gamma)}l^{\alpha(1-\gamma)-1} &= \overline{\mu}_0\bar{\iota}(1/\alpha - 1)\overline{V}.
\end{aligned}
\tag{52}
$$

Substitution of the first equation into the second gives $\alpha\bar{c} = \overline{\mu}_0\bar{\iota}(1 - \alpha)l$, which gives the expression in the statement of the lemma for leisure in term of consumption. Using this together with $\mu_\theta(l) = \overline{\mu}_0(1 - l)$, the mean growth in productivity satisfies $\mu_\theta(l)\bar{\iota}(1 - \alpha) = \overline{\mu}_0\bar{\iota}(1 - \alpha) - \alpha\bar{c}$. The first-order condition for consumption implies

$$
\frac{\rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1 - \gamma} - \bar{c}(1 - \alpha)\overline{V} = \bar{c}\overline{V}\frac{\overline{\gamma}}{1 - \gamma}.
$$

Substituting the above into the Hamilton-Jacobi-Bellman equation and dividing by $\overline{V}$ gives

$$\frac{\rho}{1-\gamma} = \overline{c}\left(\frac{\overline{\gamma}}{1-\gamma} - \alpha\right) + \left(-\tau_a + (1-\tau_s)\rho + [\overline{\mu}_0(\hat{l}-1) + \tau_s\rho + \overline{\mu}_0]\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2\right)(1-\alpha).$$

This rearranges to

$$\rho = \overline{c}\gamma + \left(-\tau_a + (1-\tau_s)\rho + [\overline{\mu}_0\hat{l} + \tau_s\rho]\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2\right)(1-\overline{\gamma})$$

where I used $\overline{\gamma} - \alpha(1-\gamma) = \gamma$. Rearrangement gives the expression for consumption. $\qquad\square$

Proof of Lemma 4.2. The equation $l^*(\overline{\iota}, \hat{l}) = \hat{l}$ rearranges to

$$\gamma\overline{\mu}_0\overline{\iota}(1/\alpha - 1)\hat{l} = \rho - \left([\overline{\mu}_0\hat{l} + \tau_s\rho]\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2 - \tau_a + (1-\tau_s)\rho\right)(1-\overline{\gamma}),$$

which obviously has a unique solution that may be found by rearranging,

$$\gamma\overline{\mu}_0\overline{\iota}(1/\alpha - 1)\hat{l} + \overline{\mu}_0\hat{l}\overline{\iota}(1-\overline{\gamma}) = \rho + \left(\tau_s\rho\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2 - \tau_a + (1-\tau_s)\rho\right)(\overline{\gamma}-1)$$

which simplifies to

$$(1/\alpha - 1)\overline{\mu}_0\overline{\iota}\hat{l} = \frac{1}{\overline{\gamma}}\left[\rho + \left(\tau_s\rho\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2 - \tau_a + (1-\tau_s)\rho\right)(\overline{\gamma}-1)\right] = \rho\tilde{x}(\overline{\iota};\tau)$$

where I used $\gamma(1/\alpha - 1) + 1 - \overline{\gamma} = (\gamma + \alpha(1-\gamma))(1/\alpha - 1) = \overline{\gamma}(1/\alpha - 1)$. It follows that $\overline{\iota}/\tilde{x}(\overline{\iota};\tau) = E(\hat{l}) = (\rho/\overline{\mu}_0)\alpha(1-\alpha)^{-1}\hat{l}^{-1}$, which gives the expression for leisure. The expression for consumption follows from this together with the relationship $\alpha\overline{c} = \overline{\mu}_0\overline{\iota}(1-\alpha)l$ derived in Lemma 4.1. To verify the form of the value function, we use the first-order condition for consumption in the proof of Lemma 4.1 to obtain $\overline{V} = \rho\overline{c}^{(1-\alpha)(1-\gamma)-1}l^{\alpha(1-\gamma)}$ and substitute the above expressions for leisure and consumption. $\qquad\square$

D.3 Transfers and debt policy

In this section I specify the transfers and debt policy necessary to support the optimal linear taxes characterized in Proposition 4.5. First note that

$$\overline{Z} = (1-\beta)Z\left(\frac{\rho_D\eta_E}{\rho_D - \mu_\theta(l)}\right)^{-\beta}\overline{L}^\beta = \frac{(\rho_D - \mu_\theta(l))}{\eta_E\rho_D}(1-\beta)Y \tag{53}$$

where Y denotes aggregate output. The utility that agents obtain is then a solution to

$$Y = Z\left(\frac{\rho_D\eta_E}{\rho_D - \mu_\theta(l)}\right)^{1-\beta}\overline{L}^\beta = \left(\frac{\rho_D\eta_E\overline{c}_r(l)}{\rho_D - \mu_c(l)} + 1 - \eta_E\right)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}} =: \kappa(l)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}.$$

The right-hand side of the above is the sum of the stationary level of worker consumption and entrepreneur consumption when the latter begins at $\overline{c}_r(l)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}$ and grows at μ_c. Steady-state utility at birth in consumption units is $[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}} = Y/\kappa(l)$ and so by Lemma E.3 the initial wealth of entrepreneurs must be

$$a_E = \rho^{-1}[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}x(l)^{\frac{\overline{\gamma}}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}} = \rho^{-1}Yx(l)^{\frac{\overline{\gamma}}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}\kappa(l)^{-1}.$$

54

Further, we have by definition $\bar{c}_r(l) = x(l)^{\frac{1}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}$. Consequently, we can write

$$\kappa(l) = \frac{\rho_D \eta_E x(l)^{\frac{1}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}}{\rho_D - (1-\bar{\gamma})\sigma^2 E(l)^2 x(l)^2/2} + 1 - \eta_E. \tag{54}$$

To obtain utility $Y/\kappa(l)$ workers require $(Y/\rho)\kappa(l)^{-1}$ units of wealth. They obtain $\beta Y/\rho$ from their labor income, and so the value of transfers must be $T = (Y/\rho)(1/\kappa(l) - \beta)$. Since the entrepreneurs also earn wage income the after-tax value of their firm at birth must equal $(Y/\rho)\left(x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}/\kappa(l) - \beta\right)$. The value of the firm is $(1-\tau_d)\overline{Z}/(\rho - \mu_\theta(l))$, and so using (53) the tax on dividends is

$$\frac{(1-\tau_d)(\rho_D - \mu_\theta(l))}{(\rho - \mu_\theta(l))\eta_E \rho_D}(1-\beta)Y = \frac{Y}{\rho}\left(\frac{x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}}{\kappa(l)} - \beta\right)$$

$$\tau_d = 1 - \frac{\rho_D(\rho - \mu_\theta(l))}{\rho(\rho_D - \mu_\theta(l))}\left(\frac{x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}} - \kappa(l)}{\kappa(l)(1-\beta)} + 1\right)\eta_E.$$

The revenue from the dividends tax minus the flow transfers as a fraction of output Y is then

$$\tau_d(1-\beta) - \rho_D(1-\eta_E)\frac{T}{Y} = 1 - \beta - \frac{\rho_D(\rho - \mu_\theta(l))}{\rho(\rho_D - \mu_\theta(l))}\left(\frac{x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}}{\kappa(l)} - \beta\right)\eta_E$$

$$- \frac{\rho_D}{\rho}(1-\eta_E)\left(\frac{1}{\kappa(l)} - \beta\right).$$

The interest payments paid by the government as a fraction of profits $(1-\beta)Y$ must then be

$$1 - \frac{\rho_D}{\rho}\left[\frac{(\rho - \mu_\theta(l))}{(\rho_D - \mu_\theta(l))}\left(\frac{x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}} - \kappa(l)}{\kappa(l)(1-\beta)} + 1\right)\eta_E + (1-\eta_E)\left(\frac{1/\kappa(l) - \beta}{1-\beta}\right)\right]. \tag{55}$$

To demystify the above it is useful to consider some special cases. If $\beta = 0$ and $\eta_E = 1$, then

$$\tau_d = 1 - \frac{\rho_D(\rho - \mu_\theta(l))}{\rho(\rho_D - \mu_\theta(l))}\left(\frac{x(l)^{\frac{\bar{\gamma}}{1-\bar{\gamma}}}l^{-\frac{\alpha}{1-\alpha}} - \kappa(l)}{\kappa(l)} + 1\right) = 1 - \frac{\rho_D(\rho - \mu_\theta(l))}{\rho(\rho_D - \mu_\theta(l))}\left(\frac{\rho_D - \mu_c}{\rho_D}\right).$$

In the absence of any discounting across generations we have $\rho_D = \rho$, and hence $\tau_d = \mu_c/\rho$, and so by Corollary 4.6 the dividends tax takes the opposing sign of the taxes raised on capital and wealth. Further, if $\alpha = 0$ then we have $\kappa(\underline{l}) = 1$ and $x(\underline{l}) = 1$, and so

$$\tau_d = 1 - \frac{\rho_D(\rho - \mu_\theta(\underline{l}))}{\rho(\rho_D - \mu_\theta(\underline{l}))} = \frac{(\rho_D - \rho)\mu_\theta(\underline{l})}{\rho(\rho_D - \mu_\theta(\underline{l}))}.$$

This expression is negative, and so in this case the government is running a surplus, the proceeds of which are used to fund a subsidy to firm profits. At birth, the investors take possession of the firms (since the entrepreneurs need not retain any equity). The debt held by the government (i.e. the negative of the amount issued) as a fraction of output is

$$\frac{1}{\rho - \rho_D}\left(1 - \frac{\rho_D}{\rho}\left[\frac{(\rho - \mu_\theta(l))}{(\rho_D - \mu_\theta(l))}\right]\right) = -\frac{1}{\rho}\left(1 - \frac{\rho_D}{\rho_D - \mu_\theta(l)}\right)$$

55

In this case the value of each firm at birth is Y/ρ. To understand who holds the debt in this economy, note that the government holds $(Y/\rho)\rho_D/(\rho_D - \mu_\theta(l))$ and so the investors must hold the negative of this quantity. Since the investors own all of the firms in the economy, they earn all (after-tax) profits, and so after purchasing new firms with a flow of $\rho_D Y/\rho$ government bonds their net revenues every instant are

$$(1 - \tau_d)Y - \frac{\rho_D}{\rho}Y - \frac{Y}{\rho}\frac{\rho_D}{(\rho_D - \mu_\theta(l))}(\rho - \rho_D)$$

$$= \frac{Y}{\rho}\left[\rho - \rho_D + (\rho - \rho_D)\frac{\mu_\theta(l)}{(\rho_D - \mu_\theta(l))} - \frac{\rho_D}{(\rho_D - \mu_\theta(l))}(\rho - \rho_D)\right] = 0$$

as expected.

D.4 General expression for optimal taxes

In this appendix I derive a general expression for the optimal taxes on interest and wealth under the assumption $\overline{\mu}_0 = \rho$ maintained at the end of Section 4. Recall that under this assumption the first-order condition characterizes the optimal retention choice and solves

$$0 = \frac{[1 + \alpha(\gamma - 1)]}{(\gamma - 1)(1 - 2\alpha)}\tilde{x}'(\overline{\iota}; \tau)\overline{\iota} + \tilde{x}(\overline{\iota}; \tau).$$

In order to implement the efficient consumption volatility $\sigma_c^*(u_\mathrm{r})$, we must choose taxes such that (27) is satisfied for $\overline{\iota} = \sigma_c^*(u_\mathrm{r})/\sigma$ and that the condition in Lemma 4.4 holds. The expressions for $\tilde{x}$ and $\tilde{x}'$ are then

$$\tilde{x}(\overline{\iota}; \tau) = \frac{1}{\overline{\gamma}}\left[1 + \left(\tau_s\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/[2\rho] - \tau_a/\rho + 1 - \tau_s\right)(\overline{\gamma} - 1)\right]$$

$$\tilde{x}'(\overline{\iota}; \tau) = \frac{1}{\overline{\gamma}}\left(\tau_s - \sigma^2\overline{\iota}/\rho\right)(\overline{\gamma} - 1).$$

Consequently, the quadratic in (27) becomes (multiplying by $\rho\overline{\gamma}$)

$$0 = \frac{[1 + \alpha(\gamma - 1)]}{(\gamma - 1)(1 - 2\alpha)}\left(\rho\tau_s\overline{\iota} - \sigma^2\overline{\iota}^2\right)(\overline{\gamma} - 1) + \rho + \left(\rho\tau_s\overline{\iota} - \overline{\gamma}\sigma^2\overline{\iota}^2/2 - \tau_a + \rho(1 - \tau_s)\right)(\overline{\gamma} - 1).$$

Using $\tau_a = \rho\tau_s(\sigma_c/\sigma - 1) - \overline{\gamma}(1 - \overline{\gamma})\sigma_c^2$ and $\overline{\iota} = \sigma_c/\sigma$ and simplifying gives

$$0 = \frac{[1 + \alpha(\gamma - 1)]}{(\gamma - 1)(1 - 2\alpha)}\left(\rho\tau_s\sigma_c/\sigma - \sigma_c^2\right)(\overline{\gamma} - 1) + \rho$$

$$+ \left(\rho\tau_s\sigma_c/\sigma - \overline{\gamma}\sigma_c^2/2 + \rho\tau_s(1 - \sigma_c/\sigma) + \overline{\gamma}(1 - \overline{\gamma})\sigma_c^2 + \rho(1 - \tau_s)\right)(\overline{\gamma} - 1)$$

$$= \frac{[1 + \alpha(\gamma - 1)](1 - \alpha)}{1 - 2\alpha}\left(\rho\tau_s\sigma_c/\sigma - \sigma_c^2\right) + \overline{\gamma}\rho + \overline{\gamma}(1/2 - \overline{\gamma})(\overline{\gamma} - 1)\sigma_c^2$$

and simplifying then gives

$$\frac{[1 + \alpha(\gamma - 1)](1 - \alpha)}{\overline{\gamma}(1 - 2\alpha)}\left(\rho\tau_s\sigma_c/\sigma - \sigma_c^2\right) = (\overline{\gamma} - 1/2)(\overline{\gamma} - 1)\sigma_c^2 - \rho.$$

Further simplification then gives

$$\tau_s = \frac{1}{\rho\sigma_c/\sigma}\left(\sigma_c^2 + \frac{\overline{\gamma}(1-2\alpha)}{[1+\alpha(\gamma-1)](1-\alpha)}\left[(\overline{\gamma}-1/2)(\overline{\gamma}-1)\sigma_c^2 - \rho\right]\right). \tag{56}$$

E   Competitive versus signaling equilibria

As noted in the main text, an earlier draft of this paper derived optimal taxes for a different solution concept, in which the price of each entrepreneur's business was assumed to be independent of the portfolio choices of the agent. Such an allocation could be expected to prevail if the portfolio choices of the entrepreneur were unobservable. In this case, taxes on interest income are necessary for the existence of a rational expectations equilibria with nonzero effort, because if the price of the firm is equal to the present discounted value of output and savings are untaxed, then the owner has no incentive to retain ownership and investors rationally expect zero effort.

It is always possible to choose taxes so that the first-order conditions of the entrepreneur are consistent with the consumption and leisure choices in the optimal restricted-action allocations. However, this "competitive" notion of equilibrium is fragile, as these first-order conditions are not always sufficient to solve the consumer's problem and non-existence of equilibrium may arise.[22] Essentially, the possibility of varying consumption, effort and the amount of shares retained leads the maximand in the consumer problem to be non-concave, and so the first-order conditions of the consumer problem only characterize the equilibrium when agency frictions are sufficiently large to discourage "double deviations" (in which the entrepreneur sells shares and reduces effort jointly). For this reason, the equilibrium notion adopted in the main text is of a "signaling" type, in which the portfolio choice of the entrepreneur is observable to investors and affects both the price of the firm and the incentives for continued effort. For the market structure in the main text, no further assumptions are necessary to ensure existence of equilibria.

However, although the expressions for taxes (and the robustness of existence) differ across the two equilibrium notions, the amount of tax revenue raised per unit of wealth does not (whenever both allocations exist). Further, the roles played by taxes are qualitatively the same across the two allocations. For this reason I outline the competitive equilibrium taxes in this appendix.

Definition E.1. *Given an interest rate r, linear taxes $\tau = (\tau_d, \tau_s, \tau_{cg}, \tau_a)$, and expectations of*

---

[22]The possibility of non-existence was erroneously overlooked in previous drafts, and so some of the allocations did not depict true competitive equilibria.

effort exerted governed by $(\hat{l}_t)_{t \geq 0}$, *the problem of an entrepreneur with firm size* θ *and wealth* a *is*

$$V(a,\theta) = \max_{\bar{c},l,\bar{\iota}} \mathbb{E}^l \left[\int_0^\infty \rho e^{-\rho t} u(\bar{c}_t a_t, l_t) dt \right]$$

$$da_t = [(1-\tau_s)(r + \rho_D) - \tau_a - \bar{c}_t] a_t dt + \bar{\iota}_t a_t dR_t(l_t; \hat{l}_t)$$

$$d\theta_t = \mu_\theta(l_t)\theta_t dt + \sigma_\theta(l_t)\theta_t dZ_t$$

$$(a_0, \theta_0) = (a, \theta).$$

Notice that the homotheticity of preferences ensures that the sole effect of a constant linear profits tax is to scale the initial wealth of the agent state-by-state, leaving the portfolio and effort decisions unaffected. The notion of equilibrium adopted here is of a standard rational expectations type: agents optimize, markets clear, and expectations are consistent with individual incentives.

Definition E.2. *Given constant linear taxes* τ, *a competitive stock market equilibrium consists of wages* w, *effort expectations governed by* $\hat{l}$, *and policy functions* $(\bar{c}, l, \bar{\iota}) = (\bar{c}_t, l_t, \bar{\iota}_t)_{t \geq 0}$ *for consumption, leisure, and investment, such that the following hold:*

- *The policy functions* $(\bar{c}, l, \bar{\iota})$ *solve the entrepreneur's problem in Definition* E.1 *given the expectations for leisure* $\hat{l}$ *and taxes* τ.

- *The outside investors break even in expectation, or* $l_t = \hat{l}$ *for all* $t \geq 0$ *almost surely.*

- *The markets for goods and labor clear every instant.*

- *The government budget constraint is satisfied.*

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

The transfers to workers and the level of government debt will be set so that the goods market clears. The first observation relevant for our characterization is that if expectations are constant, then for for a given choice of effort, the entrepreneur's problem becomes a standard portfolio problem of Merton-Samuelson type and admits a homogeneous solution. However, for an arbitrary set of taxes and prices, the first-order conditions need not characterize the solution since the drift in the law of motion for wealth is not concave in $(c, l, \bar{\iota})$. I will first proceed under the assumption that such first-order conditions characterize the solution to the individual problem, and then derive sufficient conditions to ensure that this approach is valid in terms of deep parameters of the model.

Lemma E.1. *When the first-order approach is valid and effort expectations are constant, the choice of leisure is constant, and the policy functions for consumption and investment are of the form* $c(a) = \bar{c}a$ *and* $\iota(a) = \bar{\iota}a$ *for some constants* $\bar{c}$ *and* $\bar{\iota}$, *where* $(\bar{c}, l)$ *solve*

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

$$\frac{\rho - \overline{\gamma c}}{1 - \bar{\gamma}} = (1-\tau_s)\rho - \tau_a + \frac{((1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s \rho)^2}{2\bar{\gamma}\sigma^2(1-\tau_{cg})^2}$$

(57)

and investment is given by

$$\bar{\iota} = \frac{(1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s\rho}{\bar{\gamma}\sigma^2(1 - \tau_{cg})^2}.$$

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

$$\rho V(a) = \max_{\bar{c},l,\bar{\iota}} \rho \frac{((\bar{c}a)^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\gamma} + [-\tau_a - \bar{c} + (1 - \tau_s)(r + \rho_D)]aV'(a)$$

$$+ \left[(1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D)\right]\bar{\iota}aV'(a) + \frac{\sigma^2a^2}{2}\bar{\iota}^2(1 - \tau_{cg})^2V''(a). \tag{58}$$

We now assume a solution to this equation of the form $V(a) = \overline{V}(\hat{l})a^{1-\bar{\gamma}}/(1-\gamma)$ for some $\overline{V}(\hat{l})$, so that $aV'(a) = (1-\alpha)\overline{V}(\hat{l})a^{1-\bar{\gamma}}$ and $a^2V''(a) = -\bar{\gamma}(1-\alpha)\overline{V}(\hat{l})a^{1-\bar{\gamma}}$. Substitution gives

$$\frac{\rho\overline{V}(\hat{l})}{1-\gamma} = \max_{\bar{c},l,\bar{\iota}} \frac{\rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\gamma} + [-\tau_a - \bar{c} + (1 - \tau_s)(r + \rho_D)](1-\alpha)\overline{V}(\hat{l})$$

$$+ \left(\left[(1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D)\right]\bar{\iota} - \frac{\bar{\gamma}\sigma^2}{2}(1 - \tau_{cg})^2\bar{\iota}^2\right)(1-\alpha)\overline{V}(\hat{l}). \tag{59}$$

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

$$(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma} = \frac{\bar{c}}{\rho}\overline{V}(\hat{l})$$

$$(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma} = \frac{l}{\rho\alpha}(\bar{\mu}_0 - \bar{\mu}_1)(1-\alpha)(1 - \tau_{cg})\bar{\iota}\overline{V}(\hat{l}) = \frac{1}{E(l)}(1 - \tau_{cg})\bar{\iota}\overline{V}(\hat{l})$$

$$\bar{\iota} = \frac{(1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D)}{\bar{\gamma}\sigma^2(1 - \tau_{cg})^2}.$$

Substituting the third into the second and using the first gives

$$\frac{\bar{c}}{\rho} = \frac{(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\overline{V}(\hat{l})} = \frac{(1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D)}{E(l)\bar{\gamma}\sigma^2(1 - \tau_{cg})}, \tag{60}$$

which is the first equation in (57). Substituting into the Hamilton-Jacobi-Bellman equation and dividing by $\overline{V}(\hat{l})/(1-\gamma)$ gives

$$\rho = \frac{\rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\overline{V}(\hat{l})} + \left((1 - \tau_s)(r + \rho_D) - \tau_a - \bar{c} + \frac{((1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D))^2}{2\bar{\gamma}\sigma^2(1 - \tau_{cg})^2}\right)(1 - \bar{\gamma}).$$

We then have $\overline{V}(\hat{l}) = \rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}/\bar{c}$, where $\bar{c}$ and l solve (60) and

$$\rho = \bar{\gamma}\bar{c} + \left((1 - \tau_s)(r + \rho_D) - \tau_a + \frac{((1 - \tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r + \rho_D))^2}{2\bar{\gamma}\sigma^2(1 - \tau_{cg})^2}\right)(1 - \bar{\gamma})$$

which rearranges to the second equation in (57) when $r = \rho_S$. $\qquad\square$

59

If we impose $\bar{l} = l$, then Lemma E.1 completely determines the law of motion of the agent's wealth and consumption, given the taxes on income and wealth. Since the agent faces a portfolio problem of Merton-Samuelson type, the fraction of wealth invested is the expected excess return on one's business divided by the volatility times the degree of risk aversion, or

$$\bar{\iota} = \frac{\tau_s \rho - \tau_{cg} \mu_\theta(l)}{\overline{\gamma} \sigma^2 (1 - \tau_{cg})^2}. \tag{61}$$

For simplicity, in the following I will assume common taxes on capital income. Although this is not the only choice possible, it illustrates the distinct roles played by taxes on the flow of income versus the stock of wealth, and simplifies the following expressions.

Proposition E.2. *For sufficiently high values of σ, the first-order approach is valid at the optimal allocation with linear taxes when η_E is sufficiently low, and the allocation may be implemented with common taxes on capital income $\tau_k := \tau_s = \tau_{cg}$ and wealth τ_a, where*

$$\tau_k = \frac{\overline{\gamma} \sigma^2 E(l_r^*) x(l_r^*)}{\rho - \mu_\theta(l_r^*) + \overline{\gamma} \sigma^2 E(l_r^*) x(l_r^*)}$$
$$\tau_a = \overline{\gamma}^2 \sigma^2 E(l_r^*)^2 x(l_r^*)^2 - \rho \tau_k,$$

together with linear taxes on firm profits, lump-sum transfers to workers and government debt issuance that are chosen so that all agents obtain the utility given in Proposition 3.4.

Proof. I will first characterize the taxes, before turning to the verification of the case of validity of the first-order approach. The mean and volatility of consumption growth in the optimal restricted-action allocation are $(1 - \overline{\gamma}) \sigma^2 E(l_r^*)^2 x(l_r^*)^2 / 2$ and $\sigma E(l_r^*) x(l_r^*)$, respectively. Using the fact that the variance of consumption growth coincides with the variance of wealth growth and is given by $\sigma(1 - \tau_k) \bar{\iota}$, we use Lemma E.1 with $\hat{l} = l$ to see that the equilibrium volatility of consumption is

$$\sigma_a = \frac{\tau_k (\rho - \mu_\theta(l_r^*))}{(1 - \tau_k) \sigma \overline{\gamma}}.$$

Equating this expression with the efficient analogue gives the tax on capital income. For this tax the policy function in Lemma E.1 becomes

$$\overline{c} = \frac{\rho \tau_k (\rho - \mu_\theta(l_r^*))}{\sigma^2 \overline{\gamma} (1 - \tau_k) E(l_r^*)} = \frac{\rho \sigma_a}{\sigma E(l_r^*)} = \rho x(l_r^*).$$

By Lemma E.1, for this capital tax the drift in wealth (and hence consumption) in the competitive equilibrium allocation is given by

$$\mu_a = (1 - \tau_k) \rho - \tau_a - \overline{c} + \bar{\iota} \tau_k (\rho - \mu_\theta(l_r^*)) = (1 - \tau_k) \rho - \tau_a - \rho x(l_r^*) + \overline{\gamma} \sigma_a^2.$$

Equating this with the drift in consumption in the optimal restricted-action allocation and using the defining equality for $x(l_r^*)$ once more gives

$$\tau_a = -\rho x(l_r^*) - (1 - \overline{\gamma}) \sigma^2 E(l_r^*)^2 x(l_r^*)^2 / 2 + (1 - \tau_k) \rho + \overline{\gamma} \sigma^2 E(l_r^*)^2 x(l_r^*)^2$$
$$= ((\overline{\gamma} - 1)(\overline{\gamma} - 1/2) - (1 - \overline{\gamma})/2 + \overline{\gamma}) \sigma^2 E(l_r^*)^2 x(l_r^*)^2 - \tau_k \rho$$

60

which simplifies to the claimed wealth tax. $\qquad\square$

I now establish conditions necessary for the validity of the first-order approach.

Lemma E.3. *The utility in consumption units in the restricted-action case with optimally chosen linear taxes is* $\overline{V}^{\frac{1}{1-\overline{\gamma}}} a = \rho x(l)^{-\frac{\overline{\gamma}}{1-\overline{\gamma}}} l^{\frac{\alpha}{1-\alpha}} a$, *the policy function for consumption is* $c(a) = \rho x(l)a$, *and the policy function for investment is* $\overline{\iota} = \left(1 + \overline{\gamma}\sigma^2 E(l)x(l)/(\rho - \mu_\theta(l))\right)E(l)x(l)$.

Proof. If $l = \hat{l}$ and $\tau_s = \tau_{cg} = \tau_k$ is given in Proposition E.2 then (57) becomes

$$\frac{\overline{c}}{\rho} = \frac{\tau_k(\rho - \mu_\theta(l_r^*))}{E(l_r^*)\overline{\gamma}\sigma^2(1 - \tau_k)} = \frac{\sigma_a}{E(l_r^*)\sigma} = x(l_r^*) \tag{62}$$

which gives the coefficient of the consumption function. Substituting into the first-order condition again then gives $\overline{V} = (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}[\rho/\overline{c}] = \rho l^{\alpha(1-\gamma)}\overline{c}^{-\overline{\gamma}}$, which rearranges to the claimed expression for utility. Finally, using Proposition E.2 once more implies

$$\frac{1}{1 - \tau_k} = \frac{\rho - \mu_\theta(l) + \overline{\gamma}\sigma^2 E(l_r^*)x(l)}{\rho - \mu_\theta(l)}.$$

Substituting into (61) then gives

$$\overline{\iota} = \frac{\tau_k(\rho - \mu_\theta(l))}{(1 - \tau_k)^2\sigma^2\overline{\gamma}} = \frac{(\rho - \mu_\theta(l) + \overline{\gamma}\sigma^2 E(l)x(l))}{(\rho - \mu_\theta(l))}\frac{(\rho - \mu_\theta(l))}{\sigma^2\overline{\gamma}}\frac{\overline{\gamma}\sigma^2 E(l)x(l)}{(\rho - \mu_\theta(l))}$$

which simplifies as claimed. $\qquad\square$

The following shows that Corollary 4.6 continues to hold in this different market structure.

Corollary E.4. *The revenue raised from the taxes on interest, wealth and capital gains per unit of wealth is* $\overline{\gamma}(\overline{\gamma} - 1)\sigma^2 E(l_r^*)^2 x(l_r^*)^2$. *Further, if* $\gamma > 1$ *this quantity is increasing in the number of workers per entrepreneur.*

Proof. Using Lemma E.1 and Proposition E.2, the taxes raised on interest, wealth, and capital gains are $\tau_a + \rho\tau_k + \tau_k(\mu_\theta(l_r^*) - \rho)\overline{\iota} = \overline{\gamma}^2\sigma^2 E(l_r^*)^2 x(l_r^*)^2 - \overline{\gamma}\sigma^2 E(l_r^*)^2 x(l_r^*)^2$, which simplifies to the desired expression. The second claim then follows from Proposition 3.4, together with the fact that $(\overline{\gamma} - 1)\sigma^2 E(l)^2 x(l)^2 = \rho(1 - x(l))/(\overline{\gamma} - 1/2)$ decreases with l. $\qquad\square$

I now establish sufficient conditions for the first-order conditions in the agent's problem to be sufficient for global optimality when taxes are chosen as in Proposition E.2.

Lemma E.5 (Sufficient conditions for validity of the first-order approach)**.** *The effort level* $\hat{l}$ *will be a global maximum provided*

$$\frac{\alpha}{1 - \alpha}(\overline{\gamma} - 1)(\overline{\gamma} - 1/2)\frac{(1 - \hat{l})^4}{4\overline{\gamma}^2 K(\hat{l}, \overline{\gamma})^2} + \frac{\hat{l}(1 - \hat{l})^2}{2\overline{\gamma}K(\hat{l}, \overline{\gamma})} < \frac{\rho\alpha\sigma^2}{(1 - \alpha)(\overline{\mu}_0 - \overline{\mu}_1)^2} \tag{63}$$

where the function K *is defined by*

$$K(\hat{l}, \overline{\gamma}) = 1 - \hat{l} + \hat{l}\frac{(\hat{l}^{\alpha(\gamma - 1)/\overline{\gamma}} - 1)}{\alpha(\gamma - 1)/\overline{\gamma}}. \tag{64}$$

61

Proof. Lemma E.3 implies that the value function is given by $V(a) = \overline{V}(\hat{l})a^{1-\overline{\gamma}}/(1-\gamma)$, where $\overline{V}(\hat{l}) = \rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}$. Using this expression, the right-hand side of Hamilton-Jacobi-Bellman equation (59) becomes

$$
\max_{\overline{c},l,\overline{\iota}} \frac{\rho(\overline{c}^{1-\alpha}l^{\alpha})^{1-\gamma}}{1-\gamma} + [-\tau_a - \overline{c} + (1-\tau_s)\rho](1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}
$$
$$
+ \left(\left[(1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s\rho\right]\overline{\iota} - \frac{\overline{\gamma}\sigma^2}{2}(1-\tau_{cg})^2\overline{\iota}^2 \right)(1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}.
\tag{65}
$$

Using the expresions for taxes in Proposition E.2, this simplifies to

$$
\max_{\overline{c},\overline{\iota}} \rho \frac{(\overline{c}^{1-\alpha}l^{\alpha})^{1-\gamma}}{1-\gamma} + [-\overline{\gamma}^2\sigma^2 E(\hat{l})^2 x(\hat{l})^2 + \rho - \overline{c}](1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}
$$
$$
+ \left(\left[(1-\tau_k)\mu_\theta(l) - \mu_\theta(\hat{l}) + \rho\tau_k\right]\overline{\iota} - \frac{\overline{\gamma}\sigma^2}{2}\overline{\iota}^2(1-\tau_k)^2 \right)(1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}.
$$

The above problem is concave in $\overline{c}$ and $\overline{\iota}$, and so we may always use first-order conditions to characterize these quantities for a given l. The first-order condition for consumption is $\rho\overline{c}^{-\overline{\gamma}} = \rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}(\hat{l}/l)^{\alpha(1-\gamma)}$, which rearranges to $\overline{c} = \rho x(\hat{l})(\hat{l}/l)^{-\alpha(1-\gamma)/\overline{\gamma}}$. The optimal choice of $\overline{\iota}$ is

$$
\overline{\iota} = \frac{(1-\tau_k)\mu_\theta(l) - \mu_\theta(\hat{l}) + \rho\tau_k}{\overline{\gamma}\sigma^2(1-\tau_k)^2}.
$$

For a particular l, the terms on right-hand side of the Hamilton-Jacobi-Bellman equation that depend on l may be written

$$
\rho \frac{\overline{c}^{1-\overline{\gamma}}l^{\alpha(1-\gamma)}}{1-\gamma} - \overline{c}(1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}
$$
$$
+ \frac{1}{(1-\tau_k)^2}\left[(1-\tau_k)\mu_\theta(l) - \mu_\theta(\hat{l}) + \rho\tau_k\right]^2 \frac{(1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}}{2\overline{\gamma}\sigma^2}.
\tag{66}
$$

Using the expressions for taxes, we have

$$
\frac{1}{(1-\tau_k)}[(1-\tau_k)\mu_\theta(l) - \mu_\theta(\hat{l}) + \rho\tau_k] = \mu_\theta(l) + \frac{\rho\tau_k}{1-\tau_k} - \frac{\mu_\theta(\hat{l})}{1-\tau_k}
$$
$$
= \mu_\theta(l) + \frac{\rho\overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})}{\rho - \mu_\theta(\hat{l})} - \mu_\theta(\hat{l})\frac{(\rho - \mu_\theta(l) + \overline{\gamma}\sigma^2 E(l_r^*)x(l))}{\rho - \mu_\theta(l)}
$$
$$
= \mu_\theta(l) - \mu_\theta(\hat{l}) + \overline{\gamma}\sigma^2 E(l_r^*)x(l).
$$

Using this last expression and the optimal consumption value, (66) becomes

$$
\rho \left((\hat{l}/l)^{-\alpha(1-\gamma)(1-\overline{\gamma})/\overline{\gamma}}\frac{l^{\alpha(1-\gamma)}}{1-\overline{\gamma}} - (\hat{l}/l)^{-\alpha(1-\gamma)/\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)} \right)(1-\alpha)[\rho x(\hat{l})]^{1-\overline{\gamma}}
$$
$$
+ \left[\mu_\theta(l) - \mu_\theta(\hat{l}) + \overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})\right]^2 \frac{(1-\alpha)\rho^{1-\overline{\gamma}}x(\hat{l})^{-\overline{\gamma}}\hat{l}^{\alpha(1-\gamma)}}{2\overline{\gamma}\sigma^2}.
\tag{67}
$$

The first term in parentheses simplifies to

$$\frac{1}{1-\overline{\gamma}}\hat{l}^{-\alpha(1-\gamma)(1/\overline{\gamma}-1)}l^{\alpha(1-\gamma)/\overline{\gamma}} - \hat{l}^{-\alpha(1-\gamma)(1/\overline{\gamma}-1)}l^{\alpha(1-\gamma)/\overline{\gamma}} = \frac{\overline{\gamma}}{1-\overline{\gamma}}\hat{l}^{-\alpha(1-\gamma)(1/\overline{\gamma}-1)}l^{\alpha(1-\gamma)/\overline{\gamma}}.$$

Maximizing (67) is then equivalent to maximizing

$$G(l,\hat{l}) = \frac{\rho x(\hat{l})\overline{\gamma}}{1-\overline{\gamma}}(l/\hat{l})^{\alpha(1-\gamma)/\overline{\gamma}} + \left[\mu_\theta(l) - \mu_\theta(\hat{l}) + \overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})\right]^2 \frac{1}{2\overline{\gamma}\sigma^2}. \tag{68}$$

The first-order conditions will characterize the solution provided $G(\hat{l},\hat{l}) > G(1,\hat{l})$. This inequality is equivalent to

$$\frac{\rho x(\hat{l})\overline{\gamma}}{1-\overline{\gamma}}[1 - \hat{l}^{-\alpha(1-\gamma)/\overline{\gamma}}] + \frac{1}{2\overline{\gamma}\sigma^2}\left(\left[\overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})\right]^2 - \left(\left[\overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l}) + \mu_\theta(1) - \mu_\theta(\hat{l})\right]^+\right)^2\right) > 0. \tag{69}$$

The term in parentheses is at least as large as $-2\overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})(\mu_\theta(1) - \mu_\theta(\hat{l})) - (\mu_\theta(1) - \mu_\theta(\hat{l}))^2$, and so using $\mu_\theta(1) - \mu_\theta(\hat{l}) = -(\overline{\mu}_0 - \overline{\mu}_1)(1-\hat{l})$ and multiplying by $2\overline{\gamma}\sigma^2$, the desired inequality is exactly

$$2\overline{\gamma}\sigma^2 E(\hat{l})x(\hat{l})(\overline{\mu}_0 - \overline{\mu}_1)(1-\hat{l}) > 2\overline{\gamma}\sigma^2\frac{\rho x(\hat{l})\overline{\gamma}}{1-\overline{\gamma}}[\hat{l}^{\alpha(\gamma-1)/\overline{\gamma}} - 1] + (\overline{\mu}_0 - \overline{\mu}_1)^2(1-\hat{l})^2. \tag{70}$$

Rearranging (70) and using $E(\hat{l})(\overline{\mu}_0 - \overline{\mu}_1) = \rho\alpha/[(1-\alpha)\hat{l}]$ gives

$$(\overline{\mu}_0 - \overline{\mu}_1)^2(1-\hat{l})^2 < 2E(\hat{l})(\overline{\mu}_0 - \overline{\mu}_1)\left(1 - \hat{l} + \frac{\overline{\gamma}\hat{l}}{\alpha(\gamma-1)}[\hat{l}^{\alpha(\gamma-1)/\overline{\gamma}} - 1]\right)\overline{\gamma}\sigma^2 x(\hat{l},\overline{\gamma})$$

$$:= 2E(\hat{l})(\overline{\mu}_0 - \overline{\mu}_1)K(\hat{l},\overline{\gamma})\overline{\gamma}\sigma^2 x(\hat{l},\overline{\gamma})$$

where K is defined in (64). The leisure level $l = \hat{l}$ will be a global maximum if

$$x^*(\hat{l},\overline{\gamma}) := \frac{(\overline{\mu}_0 - \overline{\mu}_1)(1-\hat{l})^2}{2\overline{\gamma}\sigma^2 E(\hat{l})K(\hat{l},\overline{\gamma})} < x(\hat{l},\overline{\gamma})$$

Using the defining quadratic for $x(\hat{l},\overline{\gamma})$, we see that this is equivalent to $0 = \sigma^2 E(l)^2(\overline{\gamma}-1)(\overline{\gamma}-1/2)x^*(\hat{l},\overline{\gamma})^2 + \rho x^*(\hat{l},\overline{\gamma}) < \rho$, or

$$\sigma^2 E(l)^2(\overline{\gamma}-1)(\overline{\gamma}-1/2)\frac{(\overline{\mu}_0 - \overline{\mu}_1)^2(1-\hat{l})^4}{4\overline{\gamma}^2\sigma^4 E(\hat{l})^2 K(\hat{l},\overline{\gamma})^2} + \rho\frac{(\overline{\mu}_0 - \overline{\mu}_1)(1-\hat{l})^2}{2\overline{\gamma}\sigma^2 E(\hat{l})K(\hat{l},\overline{\gamma})} < \rho$$

which simplifies to (63). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

F   Numerical method

In this appendix I describe the numerical method used to compute the examples given in the paper.

F.1 Principal-agent problem

I will solve the Hamilton-Jacobi-Bellman equation in Proposition A.1 using the method of Kushner and Dupuis [2001]. The idea is to approximate the solution of the continuous-time continuous-state control problem with a simpler discrete-time problem in which the state assumes only finitely-many values. I refer the reader to Kushner and Dupuis [2001] for the theory justifying the approach and only outline here the choice of approximating chains. Recall from (38) that $u_t := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}\theta_t^{-1}$ evolves according to $du_t = \mu_u u_t dt + \sigma_u u_t dZ_t$, where

$$
\begin{aligned}
\mu_u &= \rho\left(\frac{1-(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right) + (\overline{\gamma}-1)\frac{\sigma^2}{2}E(l)^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma} \\
&\quad + \frac{1}{2}(\sigma E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma} - \sigma_\theta(l))^2 - \mu_\theta(l) + \frac{\sigma_\theta^2(l)}{2} \\
\sigma_u &= \sigma E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma} - \sigma_\theta(l)
\end{aligned}
\tag{71}
$$

and $c = \overline{c}u_t\theta_t$. I will first exploit the homogeneity of the problem in order to simplify the numerical analysis. Denote the value function of the principal in terms of normalized utility and output by $\overline{V}(u,\theta)$, and note that it solves the Hamilton-Jacobi-Bellman equation

$$
\rho\overline{V} = \max_{\substack{\overline{c}\geq 0 \\ l\in[\underline{l},1]}} \theta 1_{l<1} - \overline{c}u\theta + \mu_u u\overline{V}_1 + \frac{\sigma_u^2 u^2}{2}\overline{V}_{11} + \mu_\theta\theta\overline{V}_2 + \frac{\sigma_\theta^2\theta^2}{2}\overline{V}_{22} + \sigma_u u\sigma_\theta\theta\overline{V}_{12}.
\tag{72}
$$

Since $\overline{V}(u,\theta) = v(u)\theta$ for some v, we know that $\theta\overline{V}_{12} = v'(u)\theta = \overline{V}_1$, and so $\overline{V}$ also solves

$$
\rho\overline{V} = \max_{\substack{\overline{c}\geq 0 \\ l\in[\underline{l},1]}} \theta 1_{l<1} - \overline{c}u\theta + (\mu_u + \sigma_u\sigma_\theta)u\overline{V}_1 + \frac{\sigma_u^2 u^2}{2}\overline{V}_{11} + \mu_\theta\theta\overline{V}_2 + \frac{\sigma_\theta^2\theta^2}{2}\overline{V}_{22}.
\tag{73}
$$

Solving (73) is more convenient than solving (72) because there are no cross partial derivatives that complicate the construction of approximating chains. The solution to (73) is the value function associated with a control problem with state variables (u,θ), payoff $\theta 1_{l<1} - \overline{c}u\theta$, and law of motion

$$
(du_t, d\theta_t) = \left((\mu_u + \sigma_u\sigma_\theta)u_t dt + \sigma_u u_t dZ_t^{(1)}, \mu_\theta\theta_t dt + \sigma_\theta\theta_t dZ_t^{(2)}\right)
\tag{74}
$$

where $Z_t^{(1)}$ and $Z_t^{(2)}$ are *independent* Brownian motions. We choose a grid $S_u = \{\underline{u}, \Delta_u, \ldots, \overline{u} - \Delta_u, \overline{u}\}$ for $\Delta_u := (\overline{u} - \underline{u})/N_u$ for some $N_u \geq 1$ and $\overline{u}$ sufficiently large that $l = 1$ is recommended at $u = \overline{u}$. Given $(u,\theta) \in S_u \times \mathbb{R}_+$, we suppose that the subsequent values of the Markov chain lie within the set $\{(u,\theta), (u \pm \Delta_u, \theta), (u, (1 \pm \Delta_\theta)\theta)\}$ for some Δ_θ. If we write $\mu_u + \sigma_u\sigma_\theta = \hat{\mu}_1 - \hat{\mu}_2$ where $\hat{\mu}_1, \hat{\mu}_2 \geq 0$, then the transition probabilities

$$
\begin{aligned}
p(u + \Delta_u, \theta) &= \frac{\Delta_t}{\Delta_u^2}\left(\frac{\sigma_u^2 u^2}{2} + \Delta_u\hat{\mu}_1 u\right) \\
p(u - \Delta_u, \theta) &= \frac{\Delta_t}{\Delta_u^2}\left(\frac{\sigma_u^2 u^2}{2} + \Delta_u\hat{\mu}_2 u\right) \\
p(u, (1 \pm \Delta_\theta)\theta) &= \frac{\Delta_t}{\Delta_\theta^2}\left(\frac{\sigma_\theta^2}{2} + \Delta_\theta\mu_\theta^{\pm}\right)
\end{aligned}
\tag{75}
$$

64

define a locally consistent Markov chain if Δ_t is sufficiently small. Writing $x(\bar{c}, l) = (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}$ for brevity, the expressions in (71) imply

$$
\begin{aligned}
\mu_u + \sigma_u \sigma_\theta &= \rho\left(\frac{1 - x(\bar{c}, l)}{1 - \overline{\gamma}}\right) + (\overline{\gamma} - 1)\frac{\sigma^2}{2} E(l)^2 x(\bar{c}, l)^2 + \frac{1}{2}(\sigma E(l) x(\bar{c}, l) - \sigma_\theta(l))^2 \\
&\quad - \mu_\theta(l) + \frac{\sigma_\theta(l)^2}{2} + \sigma^2 E(l) x(\bar{c}, l) - \sigma_\theta(l)^2 \\
&= \frac{\rho}{\overline{\gamma} - 1}(x(\bar{c}, l) - 1) - \mu_\theta(l) + \overline{\gamma}\frac{\sigma^2}{2} E(l)^2 x(\bar{c}, l)^2.
\end{aligned}
$$

Since $\overline{\gamma} > 1$, we can take

$$
\hat{\mu}_1 = \frac{\rho}{\overline{\gamma} - 1} x(\bar{c}, l) + \frac{\overline{\gamma}\sigma^2}{2} E(l)^2 x(\bar{c}, l)^2 \qquad\qquad \hat{\mu}_2 = \frac{\rho}{\overline{\gamma} - 1} + \mu_\theta(l).
$$

The discrete-time, discrete-state Bellman equation is then

$$
\begin{aligned}
v(u) = \max_{\bar{c} \geq 0, l \in [\underline{l}, 1]} \quad &\Delta_t(1_{l<1} - \bar{c}u) + e^{-\rho\Delta_t} v(u) \\
&+ e^{-\rho\Delta_t} p(u, (1 + \Delta_\theta)\theta)[(1 + \Delta_\theta)v(u) - v(u)] \\
&+ e^{-\rho\Delta_t} p(u, (1 - \Delta_\theta)\theta)[(1 - \Delta_\theta)v(u) - v(u)] \\
&+ e^{-\rho\Delta_t}(p(u + \Delta_u, \theta)(v(u + \Delta_u) - v(u)) + p(u - \Delta_u, \theta)(v(u - \Delta_u) - v(u))).
\end{aligned}
$$

Substituting the above expressions, dividing by Δ_t and simplifying gives

$$
\begin{aligned}
0 = -\frac{1}{\Delta_t}(1 - e^{-\rho\Delta_t})v(u) + \max_{\bar{c} \geq 0, l \in [\underline{l}, 1]} 1_{l<1} - \bar{c}u + e^{-\rho\Delta_t}\mu_\theta v(u) \\
+ e^{-\rho\Delta_t}\left(\hat{\mu}_1 u v^F - \hat{\mu}_2 u v^B + \sigma_u^2 u^2 v^{C2}/2\right)
\end{aligned}
\tag{76}
$$

where I wrote $v^{C2} = (v(u - \Delta_u) - 2v(u) + v(u + \Delta_u))/\Delta_u^2$, $v^F = (v(u + \Delta_u) - v(u))/\Delta_u$ and $v^B = (v(u) - v(u - \Delta_u))/\Delta_u$, for second central, forward and backward differences, respectively. Taking the limit $\Delta_t \to 0$ we write (76) as $0 = \max_{\bar{c} \geq 0, l \in [\underline{l}, 1]} 1_{l<1} - \bar{c}u + Tv$, where T is an $N_u \times N_u$ matrix with coefficients of the main diagonals given by

$$
\begin{aligned}
v(u + \Delta_u) \ &: \ \frac{\hat{\mu}_1 u}{\Delta_u} + \frac{\sigma_u^2 u^2}{2\Delta_u^2} \\
v(u) \ &: \ -\rho + \mu_\theta - \frac{1}{\Delta_u}(\hat{\mu}_1 + \hat{\mu}_2)u - \frac{\sigma_u^2 u^2}{\Delta_u^2} \\
v(u - \Delta_u) \ &: \ \frac{\hat{\mu}_2 u}{\Delta_u} + \frac{\sigma_u^2 u^2}{2\Delta_u^2}
\end{aligned}
\tag{77}
$$

and the maximand becomes $M = 1_{l<1} - \bar{c}u + \mu_\theta v + \hat{\mu}_1 u v^F - \hat{\mu}_2 u v^B + \sigma_u^2 u^2 v^{C2}/2$.

We must treat the case of logarithmic utility separately. Recall that in this we define normalized utility to be $u_t = \exp(U/(1 - \alpha))\theta^{-1}$. Appendix A.2 shows that $du_t/u_t = \mu_u dt + \sigma_u dZ_t$, where

$$
\begin{aligned}
\mu_u &= -\rho \ln \bar{c}_t - \frac{\rho\alpha}{1 - \alpha} \ln l_t - \mu_\theta(l_t) + \frac{\sigma^2}{2}(E(l_t) - 1_{l_t<1})^2 + \frac{\sigma^2}{2} 1_{l_t<1} \\
\sigma_u &= \sigma(E(l_t) - 1_{l_t<1})
\end{aligned}
\tag{78}
$$

and $c = \bar{c}u_t\theta_t$. The above homogeneity observation continues to hold and it suffices to solve a control problem with state variables (u, θ), flow payoff $\theta 1_{l<1} - \bar{c}u\theta$, and law of motion

$$(du_t, d\theta_t) = \left((\mu_u + \sigma_u\sigma_\theta)u_t dt + \sigma_u u_t dZ_t^{(1)}, \mu_\theta \theta_t dt + \sigma_\theta \theta_t dZ_t^{(2)}\right) \tag{79}$$

where $Z_t^{(1)}$ and $Z_t^{(2)}$ are independent Brownian motions, and we again choose a rectangular grid for (u, θ) and probabilities according to (75), for some $\hat{\mu}_1, \hat{\mu}_2 \geq 0$ with $\mu_u + \sigma_u\sigma_\theta = \hat{\mu}_1 - \hat{\mu}_2$. The expressions in (71) imply

$$\mu_u + \sigma_u\sigma_\theta = -\rho \ln \bar{c}_t - \frac{\rho\alpha}{1-\alpha} \ln l_t - \mu_\theta(l_t) + \frac{\sigma^2}{2}E(l_t)^2.$$

It will be convenient to write $\bar{c}_t = \bar{c}_t^*\nu$ for some $\nu \leq 1$ chosen such that $\bar{c}_t^* > 1$ everywhere, so that

$$\hat{\mu}_1 = -\rho \ln \nu - \frac{\rho\alpha}{1-\alpha} \ln l_t + \frac{\sigma^2}{2}E(l_t)^2 \qquad\qquad \hat{\mu}_2 = \rho \ln(\bar{c}_t/\nu) + \mu_\theta(l_t).$$

are non-negative. In practice I fix ν and check ex-post that $\hat{\mu}_1, \hat{\mu}_2 \geq 0$. Simplifying the discrete-time, discrete-state Bellman equation and taking the limit $\Delta_t \to 0$ once again gives $0 = \max_{\bar{c}\geq 0, l\in[\underline{l},1]} 1_{l<1} - \bar{c}u + Tv$, where T is an $N_u \times N_u$ matrix with coefficients of the main diagonals given by (77) and the maximand becomes $M = 1_{l<1} - \bar{c}u + \mu_\theta v + \hat{\mu}_1 uv^F - \hat{\mu}_2 uv^B + \sigma_u^2 u^2 v^{C2}/2$. The maximization is then equivalent to

$$1_{l<1} - \bar{c}u + \mu_\theta(l)[v - uv^B] + \left(-\frac{\rho\alpha}{1-\alpha}\ln l + \frac{\sigma^2}{2}E(l)^2\right)uv^F - \rho\ln\bar{c}uv^B + \frac{\sigma^2}{2}(E(l) - 1_{l<1})^2 u^2 v^{C2}.$$

F.2   Stationary distribution

In this appendix I describe how to compute the stationary distribution over (u, θ). Because both variables exhibit a fat tail, I will use the logarithmic transformations to $(w, y) := (\ln u, \ln \theta)$ as in Appendix A.2. So suppose we have the process

$$\begin{aligned}
dw_t &= \mu_w dt + \sigma_w dZ_t \\
dy_t &= \mu_y dt + \sigma_y dZ_t
\end{aligned} \tag{80}$$

where μ_w, μ_y, σ_w and σ_y are given in Appendix A.2. Note that we must transform the controls $\bar{c}$ and l so that they are functions of w rather than u. The construction of a locally consistent chain for the process in (80) is complicated by the perfect correlation between the noise terms for w and y. To address this point, I again follow Kushner and Dupuis [2001] and construct a locally consistent chain using the procedure explained in Phelan and Eslami [2022].

I consider a uniform grid over the logarithm of promised utility $S_w = \{\underline{w}+\Delta_w, \underline{w}+2\Delta_w, \ldots, \overline{w}-\Delta_w\}$, where $\Delta_w := (\overline{w} - \underline{w})/N_w$ for some $N_w \geq 1$. Note that there are $N_w - 1$ points in the grid S_w; there would be $N_w + 1$ points if I included $\underline{w}$ and $\overline{w}$. The state space for the problem is then $S_w \times S_y$, where $S_y = \{\underline{y} + \Delta_y, \ldots, \overline{y} - \Delta_y\}$ for some $\underline{y}$ and $\overline{y}$, where $\Delta_y := (\overline{y} - \underline{y})/N_y$ for some

$N_y \geq 1$. To construct a locally consistent chain, I proceed as if we were randomizing between two processes, a pure drift process and a pure diffusion process, that together sum to the law of motion in (80). I will write the transition probabilities as the sum of two processes that are locally consistent with each component, $p(X') = p^{(1)}(X') + p^{(2)}(X')$, for any $X' = (w', y') \neq X = (w, y)$, with $p(X)$ chosen such that probabilities sum to unity. The $p^{(1)}$ are chosen to be locally consistent with the drift terms given in the system (80), so that

$$
\begin{aligned}
p^{(1)}(w \pm \Delta_w, y) &= \frac{\Delta_t}{\Delta_w} \max\{0, \pm\mu_w\} \\
p^{(1)}(w, y \pm \Delta_y) &= \frac{\Delta_t}{\Delta_y} \max\{0, \pm\mu_y\}
\end{aligned}
\tag{81}
$$

together with the assumption that the boundary points are absorbing. It remains to choose the probabilities $p^{(2)}$ that form a locally consistent chain for

$$
(dw_t, dy_t) = (\sigma_w dZ_t, \sigma_y dZ_t)
\tag{82}
$$

and then randomize between this chain and the one defined by the probabilities in (81). I now fix a point (w, y) and suppose that for some integers $m_1, m_2 \geq 0$ (which can depend on the grid), the chain can either stay at the current point or transition to one of the two points $(w \pm m_1 \Delta_w, y \mp m_2 \Delta_y)$ with some (possibly grid-dependent) probability $\bar{p}$ further specified below. Note that I am using the fact that σ_w and σ_y necessarily have opposing signs, which is not assured for all parameters but is true for all of the examples computed in the paper. The local consistency requirements for the process in (82) are

$$
\begin{aligned}
2\bar{p} m_1^2 \Delta_w^2 &= \Delta_t \sigma_w^2 + o(\Delta_t) \\
-2\bar{p} m_1 m_2 \Delta_w \Delta_y &= \Delta_t \sigma_w \sigma_w + o(\Delta_t) \\
2\bar{p} m_2^2 \Delta_y^2 &= \Delta_t \sigma_y^2 + o(\Delta_t).
\end{aligned}
$$

Note that σ_y is either σ or zero, depending on w. If $\sigma_y = 0$, then I set $m_1 = m_2 = 0$. If $\sigma_y = \sigma$, then I choose $m_2 = 1$ and $\bar{p}$ such that the third condition holds

$$
\bar{p} = \frac{\Delta_t}{\Delta_y^2} \frac{\sigma^2}{2}.
\tag{83}
$$

In view of the local consistency conditions, I would like to choose m_1 such that $m_1^2 \sigma^2 / \Delta_y^2 = \sigma_w^2 / \Delta_w^2$, but there might be no such m_1. In this case, I fix $\overline{m} \geq 0$ and search for $m_1^* \leq \overline{m}$ that minimizes

$$
\left| m_1 - \frac{|\sigma_w| / \Delta_w}{\sigma / \Delta_y} \right|.
\tag{84}
$$

The integer m_1 that I choose is then the minimum of this m_1^* and the least integer such that the transitions remains on the grid, and the transition probability given by (83). As we explain

in further detail in Phelan and Eslami [2022], this three-point approximation can, in principle be inaccurate for an arbitrary control problem. However, in this example, I can ensure that the quantity in (84) vanishes in the region of the grid in which the process spends the most time by choosing Δ_y such that $|\sigma_w|/\Delta_w = \sigma/\Delta_y$ when σ_w equals the value associated with the highest effort. Figure 7 depicts the cdfs for consumption and firm size in the restricted-action allocation associated with the highest effort level computed using this method together with the associated closed-form solution. For readability I truncate below the top 1 percent. Reassuringly, the true and computed cdfs appear almost indistinguishable from on another.
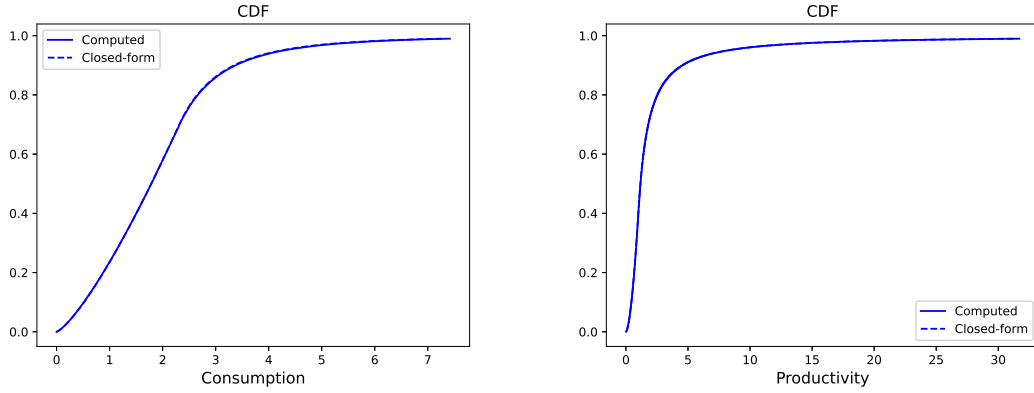

Figure 7: Computed and closed-form cdfs for restricted-action allocation.