working paper

19 26R

The Optimal Taxation of Business Owners

Thomas Phelan

FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working Paper 19-26R May 2021*

The Optimal Taxation of Business Owners

Thomas Phelan

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 optimal linear taxes on income and wealth 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 characterize optimal linear taxes when owners may save in a risk-free bond and trade shares in their businesses. The optimal utilitarian policy calls for separate taxes on firm profits, capital income, and wealth, serving distinct purposes. The tax on profits plays a redistributive role, the tax on capital income 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, human capital.

JEL Codes: D61, D63, E62.

Suggested citation: Phelan, Thomas. 2021. "The Optimal Taxation of Business Owners." Federal Reserve Bank of Cleveland, Working Paper no. 19-26r. https://doi.org/10.26509/frbc-wp-201926r.

Thomas Phelan is at the Federal Reserve Bank of Cleveland (tom.phelan@clev.frb.org). This is a revised version of the first chapter of the author's Ph.D. dissertation completed at the University of Minnesota and previously circulated as, "On the Efficient Level of Inequality in Business Income." The author has benefited from discussions with Brian Albrecht, Hengjie Ai, Anmol Bhandari, V.V. Chari, Sebastian DiTella, Sergio Ocampo, Chris Phelan, Guillaume Sublet, Randy Verbrugge, and seminar participants at various conferences.

*First version November 2019.

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 coincides with an average fall in profits of 54 percent.

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 business income were solely the return to savings, then the results of Chamley (1986) and Judd (1985) imply it ought not to be taxed at all, at least in the long run. If instead individuals were simply endowed with exogenous productivity and only choose their 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 appear to 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 provides guidance on whether we ought to levy different taxes on profits, interest income, or capital gains.

Motivated by these considerations, this paper characterizes efficient allocations and optimal 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. In order to isolate the role of private information in a partial equilibrium setting, I first alter the principal-agent model of Sannikov (2008) to incorporate persistent effects of effort on output. I assume that the output of a firm at any moment is publicly observable and independent of current choices, but that the growth of 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 future 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 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, agents typically obtain high consumption precisely because their productivity rose, and so the benefits of 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 incorporation of persistent effects of effort has important effects on the implied stationary distribution and, in particular, the thickness of the right tails of consumption and workers per entrepreneur (firm size). The tractability of the model then allows for simple comparative statics connecting changes in aggregate technology with efficient distributions of income and firm size. Specifically, in response to any change in resources or technology that increases the marginal productivity of an entrepreneur (such as an increase in workers per entrepreneur), a benevolent planner will wish to make incentives for effort more high-powered, which necessitates an increase in consumption inequality ex-post.

Following this abstract characterization I consider the implications for taxation policy. Prescriptions for taxes unavoidably depend on assumptions regarding the degree of risk-sharing present in private markets and the complexity of contracts that agents are assumed capable of signing. If agents may write contracts of arbitrary complexity with financial intermediaries, then the optimal policy calls only for lump-sum transfers between entrepreneurs and workers.[1] If the government has no advantage over the private sector at overcoming

[1] An earlier draft erroneously claimed that the optimum could be achieved with linear taxes.

agency frictions, the latter provides utility at the lowest cost possible, and the lump-sum taxes serve only to achieve the redistributive goals of the government.

However, this decentralization assumes that financial intermediaries can commit to long-term contracts and monitor the consumption of agents. Since both assumptions are unrealistic, to complement the above I characterize the optimal linear taxes on income and wealth in an environment in which agents may only save in a risk-free bond and trade shares of their firms. Within this setting the dependence of productivity growth on unobserved and non-contractible effort implies a novel effect of taxation policy: taxes on capital income affect the incentives for retaining ownership of one's business and hence the degree of risk-sharing possible 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 outside option of the owner, the return on risk-free savings. Within this linear class, the optimal policy now calls for taxes on profits, capital income, and wealth. 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, and only reduces the ex-ante value of the firm. In contrast, taxes on capital income and wealth alter the private returns to effort in one's business and consumption smoothing, respectively, and vanish as agency frictions become negligible.

Related literature. A vast literature builds upon Mirrlees (1971) to derive optimal taxes in economies with asymmetric information. Scheuer (2014) explicitly considers entrepreneurs and firm formation within a static model and allows for both 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) explore how optimal taxation policy must be altered 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 in various ways, 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 of their framework for long-run distributions of consumption or income, and because their two-period nature,[2] they cannot address how the risk borne by any agent depends on his history of productivity shocks.

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) 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 assumed to be unable to save and the evolution of each agent's productivity is identical in form to that adopted in this paper. As Jones and Kim (2018) note, when combined with stochastic death, this assumption on productivity growth can both generate

[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.

the empirically-observed 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 in the subsequent decentralization, I can investigate how the portfolio decisions of owners and the expectations of outside investors are endogenous to taxation policy.

Ai et al. (2016), Shourideh (2013), and Phelan (2019) all show how a Pareto distribution of consumption may emerge in the presence of asymmetric information with optimal contracting in private insurance markets. The difference here is the nature of the agency problem: instead of allowing entrepreneurs to divert assets to private consumption, in this paper entrepreneurs exert (unobservable) effort to improve firm productivity. I adopt the specification of random productivity governed by hidden effort in order to model the importance of individual-specific characteristics for business income. Importantly, unlike all of the above papers, the competitive equilibria of 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. 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 those papers do not explore how the degree of private risk-sharing depends on taxation policy.

For clarity, the first section of this paper analyzes a principal-agent problem between a risk-averse entrepreneur and a risk-neutral principal in partial equilibrium where flow output depends solely on their own productivity. I characterize and compute the policy functions of the principal and show numerically that typically the agent will exert more effort and bear more risk when he or she is more productive. In the subsequent analysis with a continuum of agents, output depends both on their productivity and an endogenous shadow price of labor, as the latter determines the societal benefit of an entrepreneur. This allows me to address how exogenous changes in the number of entrepreneurs, or changes in technology that primarily benefit entrepreneurs relative to workers, translate into changes in inequality both between workers and entrepreneurs and among entrepreneurs in efficient allocations. I view this as complementary to the standard approach in the optimal taxation literature, in which the path of wages (or productivity) evolves exogenously, but current income is endogenous to taxation policy. Since I model a dynamic moral hazard problem and an active stock market with endogenous beliefs, I simplify the ex-ante heterogeneity for tractability, and leave the quantitative importance of these two margins for future work.

The outline of the paper is as follows: Section 2 characterizes the optimal contract between a single agent and a principal. Section 3 extends this to an overlapping generations economy with a continuum of agents with heterogeneous ability and shows how to compute stationary distributions of income in a number of example economies. Section 4 characterizes the optimal linear taxes and Section 5 concludes. Details of the recursive techniques, numerical implementation and the welfare notions employed are relegated to the appendix.

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.

2.1 Formal setup

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

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

where the instantaneous 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 $\overline{\gamma} := 1-(1-\gamma)(1-\alpha)$ 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 preferences of the agent and so their preferences over output and consumption are

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

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 $(c_t, l_t^P)_{t\geq 0}$ and strategy $(l_t)_{t\geq 0}$, the continuation utility of the agent is

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

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 the allocation specifies consumption as a function of every finite history of output, and when choosing a strategy, the agent understands how his actions change the probability of each history and weights them accordingly. For any strategy $l = (l_t)_{t\geq 0}$ I will write $\mathbb{E}^l$ for the expectation operator associated with the output process implied by l.[3] The utility of an agent confronted with an allocation $(c_t, l_t^P)_{t\geq 0}$ when adhering to $(l_t)_{t\geq 0}$ is therefore

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

The definition of incentive compatibility is then the following.

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

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

$$V(U, \theta) = \max_{(c,l)\in\mathcal{A}^{IC}} \mathbb{E}\left[\int_0^\infty e^{-\rho t} [\theta_t 1_{l_t<1} - c_t] dt \right]$$
$$U = \rho \mathbb{E}\left[\int_0^\infty e^{-\rho t} u(c_t, l_t) dt \right]$$
$$d\theta_t = \mu_\theta(l_t)\theta_t dt + \sigma_\theta(l_t)\theta_t dZ_t, \quad \theta_0 = \theta.$$

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

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) 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 $(c_t, l_t^P)_{t\geq 0}$ 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 effort 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. Expression (7) 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 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 $\lambda V(U\lambda^{\overline{\gamma}-1}, \theta)$. $\square$

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 $v(u) := V\big(u^{1-\overline{\gamma}}/(1-\gamma), 1\big)$, 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. (2016) 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

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

For a fixed effort level the mean and volatility of utility growth depend only on $\overline{c}$. Further, for these allocations the perturbation arguments of Golosov et al. (2003) are applicable, and so the inverse Euler equation holds and we have the following sharp characterization proved in Appendix A.4.

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

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

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, $v'_{\mathrm{r}}(u;l)$ is increasing in l.

There are two important points to be taken from Proposition 2.3. First, we obviously have $v(u) \geq v_{\mathrm{r}}(u;l)$ for all $l \in [\underline{l}, 1]$, and so

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

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. Note that this sharp characterization of payoffs near zero depends on the preferences adopted and differs from Phelan and Townsend (1991) and Sannikov (2008), in which it is possible for the agent to be retired (and output to cease) at low levels of utility when utility is bounded below.

The second important consequence of Proposition 2.3 is that $v'_{\mathrm{r}}(u;l)$ 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 $v^*_{\mathrm{r}}(u) := \max_{l \in [\underline{l},1]} v_{\mathrm{r}}(u;l)$ of a restricted principal who must fix effort recommendations throughout the entirety of an agent's life, and note that Topkis' theorem implies that the associated policy function is increasing in normalized utility. Such a restricted principal therefore recommends high effort to individuals with low initial normalized utility and vice versa.

The expressions in Proposition 2.3 do not, by themselves, tell us anything about how normalized utility evolves over time and in response to shocks. However, the closed-form expression for consumption in (2.3) allows for a sharp 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* $\sigma_u = \sigma(E(l)x(l) - 1)$*, and high productivity shocks reduce normalized utility if and only if*

$$\rho < (1/\alpha - 1)(\overline{\mu}_0 - \overline{\mu}_1)l + \frac{\sigma^2}{2}(2\overline{\gamma} - 1)(\overline{\gamma} - 1). \tag{8}$$

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

$$\mu_u - \frac{\sigma_u^2}{2} = \frac{\rho}{1 - \overline{\gamma}}(1 - (\overline{c}^{1-\alpha} l^\alpha)^{1-\gamma}) + \frac{(\overline{\gamma} - 1)}{2}\sigma^2 E(l)^2 (\overline{c}^{1-\alpha} l^\alpha)^{2-2\gamma} - \mu_\theta(l) + \frac{\sigma_\theta^2(l)}{2}. \tag{9}$$

The expressions for the drift and diffusion of normalized utility follow by combining the expression (9) with the quadratic in Proposition 2.3. High productivity shocks reduce normalized utility when $\sigma_u < 0$, or $E(l)x(l) < 1$. Since the quadratic in Proposition 2.3 is increasing for positive x, rearrangement reveals that this is equivalent to (8). □

Recall that if a process $(X_t)_{t \geq 0}$ 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 $t \geq 0$. Consequently, the large-time pathwise behavior is determined not by μ_X but by the quantity $\mu_X - \sigma_X^2/2$, which may be interpreted as a kind of risk-adjusted growth rate. Lemma 2.4 therefore shows that the large-time behavior of normalized utility is governed by the difference between two terms. The first, $-\overline{\gamma}\sigma^2 E(l)^2 x(l)^2/2$, is negative and represents a force common in economies with private information, in which the principal wishes to front-load utility to relax future incentive constraints, while the second, $\mu_\theta(l) - \sigma_\theta(l)^2/2$, is simply the risk-adjusted growth rate 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 have high utility only 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. Indeed, expression (8) in Lemma 2.4 shows that in the restricted-action allocations, the dynamics of risk-sharing may be entirely reversed, in the sense that high shocks *reduce* normalized utility, and so lead the principal to recommend higher effort and more risk in the agent's utility. Further, expression (8) also shows that for sufficiently high risk-aversion or uncertainty, or a sufficiently patient agent, this will be true regardless of the recommended leisure level.

Numerical illustration. Since none of the above restricted-action value functions are globally optimal, the value function must be calculated numerically. Appendix D outlines the numerical method I use to solve the principal's problem. Recall that the complete list of parameters is $\alpha, \gamma, \rho, \sigma, \overline{\mu}_0, \overline{\mu}_1$ and $\underline{l}$. For preferences I choose $(\rho, \alpha, \overline{\gamma}) = (0.07, 0.3, 2)$, all of which are standard. The choice of the remaining parameters is more difficult, both due to uncertainty regarding the risk profiles of business incomes and the absence of an obvious way to discipline $\underline{l}$. I follow Jones and Kim (2018) and set $\sigma = 0.15$, but since I abstract from aggregate growth in Section 3, I choose lower values for growth to ensure stationarity. For simplicity I choose $(\overline{\mu}_0, \overline{\mu}_1, \underline{l}) = (0.07, 0.0, 0.5)$, which corresponds to expected productivity growth of 3.5 percent with high effort and zero growth with no effort.

Figure 1 plots the true value function alongside several restricted-action value functions. Obviously, the true value function must everywhere lie above each of these restricted-action value functions. Further, as implied by the above discussion, for low values of normalized utility the true value function is approximately equal to the restricted-action value function associated with the lowest leisure level. Figure 2 plots the policy functions for leisure and consumption together with the analogous optimal restricted-action functions. As expected, the policy functions for leisure and consumption are both increasing in normalized utility. This shows that as the amount owed to the agent increases, the principal chooses to reward the agent with leisure rather than consumption and reduces the risk associated with their

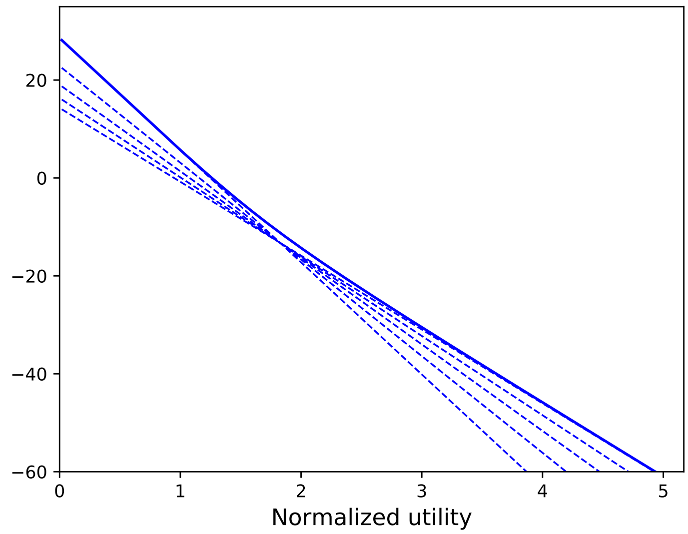

Figure 1: Efficient and restricted-action value function.

future utility.

Finally, Figure 3 plots the risk-adjusted growth in normalized utility and the diffusion of normalized utility in both the efficient and optimal restricted-action cases. As implied by Lemma 2.4, the volatility of normalized utility is everywhere negative in the restricted-action allocations, and this remains the case (and is even more pronounced) in the true efficient allocation. When combined with the monotonicity of the leisure function, this implies that high realizations of shocks lead the principal to recommend higher effort, and hence higher risk associated with future utility. Note that in all cases the efficient functions inherit the qualitative features of their restricted-action counterparts, but exhibit greater smoothness.

Before turning to the setting with a continuum of agents, it is instructive to compare the above optimal contract with those derived in similar environments with hidden actions, and with the law of motion of income that obtains in economies with exogenously incomplete markets. This will also serve as a prelude to why thick right tails of consumption and firm size emerge in this environment and highlight the novel forces present in this model. Shourideh (2013) and Phelan (2019) consider economies with idiosyncratic capital risk and

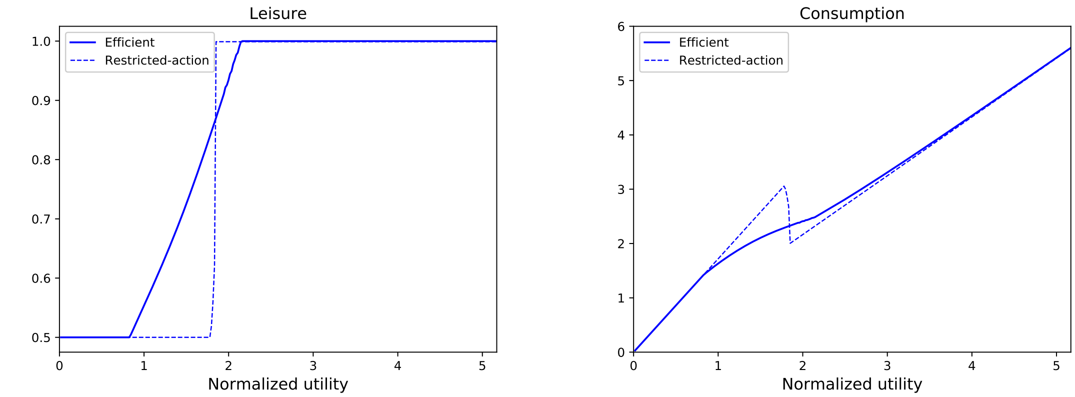

Figure 2: Efficient and restricted-action leisure and consumption function.

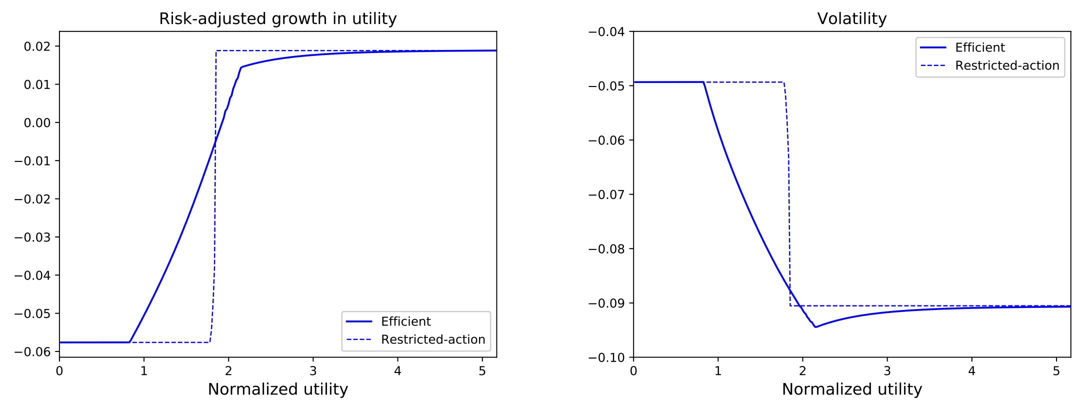

Figure 3: Efficient and restricted-action law of motion for utility.

hidden consumption. To preserve incentives for investment, the risk borne by entrepreneurs must scale with the benefits of diverting capital to private consumption. As a result, rich agents with low marginal utility of consumption also control high amounts of capital, and so the net effect on incentives to deviate is, in general, ambiguous. With homothetic preferences, these forces cancel out so that the elasticity of consumption with respect to output is common across agents, which leads Shourideh (2013) and Phelan (2019) to generate thick right (Pareto) tails. However, although they generate heterogeneous capital income and a thick tail of the wealth distribution, the mechanism adopted in these two papers cannot speak to the importance of owner characteristics mentioned in the introduction.

It is also worth contrasting the lessons drawn from this optimal contracting environment with the exogenously incomplete-markets allocation in Jones and Kim (2018). In their model entrepreneurs are unable to save, and so consumption, income, and firm profits coincide. This differs sharply from the allocations in 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 returns to effort and 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 much thicker than that for consumption.

Proposition 3.6 below formalizes this discussion 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 (for sufficiently high productivity growth), the desire to smooth consumption over time sharply restricts the efficient level of consumption growth and places tight bounds on the tail of the ensuing distribution. Before proceeding to the general equilibrium context, note 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. This allows me to derive the effects of private information on the long run distributions of consumption and firm size. 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.

3.1 Environment

Time is again continuous and indefinite. At any instant there is a continuum of agents alive in the economy with subjective discount factor ρ_S who die at rate ρ_D. A flow of ρ_D agents are born every unit of time so that the total population is fixed at unity. The preferences of agents continue to be of the form given in (1) and (2) with discount rate now $\rho := \rho_S + \rho_D$. Every agent alive at the initial date is indexed by a single variable $v \in \mathbb{R}$ identified with promised utility. To each v-agent there is an associated process $Z^v = (Z_t^v)_{t\geq 0}$ distributed according to standard Brownian motion and referred to as the noise process for agent v. These processes are independent of one another, and so by a law of large numbers for a continuum of agents,[4] the ex-post distribution of shocks across agents will coincide with the ex-ante distribution faced by a single agent. The noise processes of agents of subsequent generations will be indexed by agents' dates of birth rather than promised utility and so agents are only distinguished by date of birth and possibly type.

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 to

[4] Subject to the usual measurability caveats.

improve his productivity through effort, and have a common productivity $\theta_0 = 1$ at birth that evolves as in Section 2, $d\theta_t^v = \mu_\theta(l_t)\theta_t^v dt + \sigma_\theta(l_t)\theta_t^v dZ_t^v$. In Section 2 the flow output of an agent coincided with his productivity. In contrast, in this section the output of an entrepreneur is a function of both their productivity and the total effective labor assigned to him. To ensure that output is finite in the stationary distribution I will also suppose that $\rho_D > \mu_\theta(l)$ for all leisure levels. All agents inelastically supply $\overline{L}$ units of effective labor per unit of time, irrespective of any effort employed to improve their productivity, and so I omit labor supply from the definition of an allocation.

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 is $F(\theta, L) = Z\theta^{1-\beta}L^\beta$ per unit of time. An allocation must specify the consumption, effort exerted, and labor assigned to every member of the initial generation as a function of initial promised utility, type, and history of output, together with analogous quantities for all subsequent generations as functions of birth date and type. In the following $L_t^{v,\theta,i}$ and $L_t^{T,i}$ refer to the effective labor assigned *to* a given type (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 $(c_t^{v,\theta,i}, l_t^{v,\theta,i}, L_t^{v,\theta,i})_{t\geq 0},\ (v,\theta,i) \in \text{supp}(\Phi)$ 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.

I will denote the set of all allocations by $\mathcal{A}$, and for any $T \geq 0$ write $\mathcal{A}_T$ for the subset of allocations associated with agents born at time T. I will denote aggregate consumption, labor assignments, output, and (flow) utility by C_t, L_t, Y_t and U_t, respectively.[5]

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 $\mathcal{A}^{RF}$.

Definition 3.3. Given an initial distribution Φ over utility, productivity, and types, an allocation A satisfies promise-keeping if $U(c^{v,\theta,i}, l^{v,\theta,i}) = v$ for all $(v,\theta,i) \in \text{supp}(\Phi)$.

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

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

Definition 3.4. 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 $\mathcal{A}^{IF} := \mathcal{A}^{RF} \bigcap \mathcal{A}^{IC}$.

Agents have preferences solely over their own consumption and effort and there is no altruism across generations. For simplicity I will suppose that the planner only cares about workers' utility and places a weight of $\alpha(T) = e^{-\rho_S T}$ on an agent born at time T. This implies that the planner values the utility of an agent at any given date the same regardless of the agent's date of birth and is equivalent to the social welfare function $U^P(A) = \int_0^\infty e^{-\rho_S t} U_t dt$ where U_t is the total flow utility of workers at $t \geq 0$.[6] I now relate the planner's problem to the principal-agent problems of Section 2.

Definition 3.5. 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, and first consider the simpler problem of a planner who may trade goods and labor intertemporally at the subjective rate of time preference.

[6] To aid the reader, elaboration of the social welfare function is given in Appendix B.1.

Definition 3.6. Given an initial distribution Φ, the relaxed problem of the planner is

$$V^R(\Phi) = \max_{A \in \mathcal{A}^{IC}} \int_0^\infty e^{-\rho s t} U_t dt$$
$$\int_0^\infty e^{-\rho s t}(C_t - Y_t) dt \leq 0$$
$$\int_0^\infty e^{-\rho s t}(L_t - \overline{L}) dt \leq 0.$$

Notice that if we solve the relaxed problem and all implied distributions are constant over time, then we have solved the planner's problem beginning at that distribution. The relaxed problem still takes a distribution as an argument, but there are now only two constraints and so the interdependence across agents is captured by just two Lagrange multipliers. For each choice of multipliers, solving the above amounts to maximizing the components of the integral relevant to each generation in isolation. I will refer to this latter problem as a *generational* planner's problem.[7] The only differences between the generational planner's problem and the principal-agent problem analyzed in Section 2 are the presence of an additional constraint requiring the utility of an entrepreneur be sufficiently high to ensure truthful revelation, and the additional (static) task of assigning workers to entrepreneurs.

A generational planner facing a population of newborns must internalize the effect that the effort levels recommended to entrepreneurs have on the shadow price of labor. At any moment the assignment of workers to an entrepreneur depends solely on the shadow price of labor and productivity and solves the static problem $\max_{L \geq 0} \; Z\theta^{1-\beta}L^\beta - \lambda_L L$, 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 $\overline{Z}(\lambda_L) := (1-\beta)Z^{\frac{1}{1-\beta}}(\beta/\lambda_L)^{\frac{\beta}{1-\beta}}$.*

For any multiplier there will be a stationary density of normalized utility that depends only on the policy functions associated with unitary productivity.

[7]These are similar to the component planning problems in Farhi and Werning (2007) and Atkeson and Lucas (1992).

Definition 3.7. For a given pair $\lambda := (\lambda_R, \lambda_L)$ of multipliers, denote the associated stationary distribution over $\Omega' := \mathbb{R} \times \Theta \times \{E, W\}$ by Φ_λ. The stationary form of the goods and labor resource constraints then reduce to the following pair of equations

$$\begin{aligned} 0 &= \int_{\Omega'} \mathbb{E}[c^{v,\theta,i} - F(\theta^{v,\theta,i}, L^{v,\theta,i})]\Phi_\lambda(d\omega') \\ 0 &= \overline{L} - \int_{\Omega'} \mathbb{E}[L^{v,\theta,i}]\Phi_\lambda(d\omega'). \end{aligned} \tag{10}$$

The first equation in (10) imposes the requirement that aggregate consumption equal aggregate output, while the second imposes the requirement that the total stock of effective labor coincide with the aggregate amount assigned to entrepreneurs.

Consequently, if λ satisfies (10) 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 will 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. This amounts to solving the generational planner's problem and then varying the shadow prices of goods and labor until the resource constraints hold in the associated stationary distribution. To this end, note that just as the homotheticity of preferences and exponential growth in productivity 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 3.8. 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 of (non retired) entrepreneurs is $\int_0^\infty m(u) 1_{l(u)<1} du$, while the average consumption of entrepreneurs (retired and non retired) is $\int_0^\infty c(u)m(u)du$. The following shows that this summary measure solves a version of the Kolmogorov forward equation for a single variable. For a proof of the following, see Appendix B.3.

Lemma 3.2. *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 ODE*

$$0 = -(\rho_D - \mu_\theta(u))m(u) - [(\mu_u(u) + \sigma_\theta(u)\sigma_u(u))um(u)]' + \frac{1}{2}[\sigma_u^2(u)u^2 m(u)]''. \tag{11}$$

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 from the point of view of a planner, changes in resource constraints simply 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

$$M(\overline{u}) := \int_0^\infty m_{\overline{u}}(u) 1_{l(u)<1} du \qquad\qquad C(\overline{u}) := \int_0^\infty c(u) m_{\overline{u}}(u) du. \tag{12}$$

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

$$\begin{aligned} \overline{L} &= (\text{ave. } \theta) \times (\text{no. entrepreneurs}) \times (\text{labor per } \theta) \\ &= M(\overline{u}) \times \eta_E \times (Z\beta/\lambda_L)^{\frac{1}{1-\beta}}. \end{aligned} \tag{13}$$

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

Proposition 3.3. *The stationary level of normalized utility is the solution to*

$$\frac{\eta_E}{1-\beta} M(u) = (\eta_E C(u)/u + 1 - \eta_E)u \tag{14}$$

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

Proof. From Lemma 3.1 the output of each firm is $(1-\beta)^{-1}\overline{Z}(\lambda_L)\theta$. Using (13) 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 (14) and rearranging (13) gives utility. $\square$

Proposition 3.3 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 3.4. *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 level of normalized utility is a root of H. The first two claims follow from Topkis' theorem and the fact that H is increasing in η_E and β, and the last claim follows from the fact that Z does not appear in (14). $\square$

The third claim in Corollary 3.4 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 the associated stationary distributions of consumption and firm size, as these are simply 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 formalize the discussion following Lemma 2.4 and in particular 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 optimal restricted-action policy functions. This will also serve as a prelude to the decentralization results of Section 4, where it is shown that these distributions are precisely those that may be implemented with linear taxes.

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 (unnormalized) lifetime utility. Imposing the labor resource constraint once more and simplifying gives the following analogue of Proposition 3.3 and Corollary 3.4.

Proposition 3.5. *The optimal restricted-action leisure level is $l_{\mathrm{r}}^*(u_{\mathrm{r}})$, where u_{r} solves*

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

The associated utility is $\overline{Z}u_{\mathrm{r}}$ where

$$\overline{Z} = (1-\beta) Z \left(\frac{\rho_D \eta_E}{\rho_D - \mu_\theta(l_{\mathrm{r}}^*(u_{\mathrm{r}}))}\right)^{-\beta} \overline{L}^\beta$$

and $l_{\mathrm{r}}^*(u_{\mathrm{r}})$ is increasing in both η_E and β.

It is well-known that the stationary distribution of a geometric Brownian motion $X = (X_t)_{t\geq 0}$ that is killed at a constant rate and reinjected at some point $\overline{X}$ has a stationary distribution of the *double-Pareto* form $f(x) = Ax^{\alpha_X^+ - 1}1_{x\leq \overline{X}} + Bx^{\alpha_X^- - 1}1_{x > \overline{X}}$ for some constants $A, B > 0$ and tail parameters $\alpha_X^\pm$. The expressions for consumption in the restricted-action allocations in Proposition 2.3 then lead to the following characterization.

Proposition 3.6. *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 double-Pareto. 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

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

The proof of Proposition 3.6 is contained in Appendix B.4. Proposition 3.6 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.[8] In contrast, the mean and volatility of consumption growth are determined by the nature of the agency problem, and depend on the technological parameters only insofar as the latter affect the tightness of the incentive constraints. Indeed, Proposition 3.6 shows that we have the uniform bound $\alpha_{\pm}(l) \leq -\overline{\gamma}$ for the tail parameter for consumption, independent of the nature of the agency problem. This contrasts sharply with other models with similar multiplicative growth dynamics such as Jones and Kim (2018), in which the assumed market structure forces the above two quantities to coincide.

I now illustrate these points by plotting the upper tail parameters for consumption and firm size with $\rho_D = 0.05$ and the same parameters adopted in Figure 1. Following Jones and Kim (2018), one need not view ρ_D literally as the probability of death, but rather the probability with which one ceases to be an entrepreneur, so that this choice of ρ_D amounts to assuming that agents run their business for an average of 20 years and have the subjective rate of discount $\rho_S = 0.02$. On the left-hand side of Figure 4 all parameters except for leisure coincide with Figure 1, while on the right-hand side leisure is set at the average level $l = (\underline{l} + 1)/2$ and σ varies. As suggested by the above discussion, for all parameters plotted the tail of firm size is far thicker than that of consumption. To illustrate the importance of allowing separate determinants for consumption growth and firm size growth, Figure 5 repeats the above exercise with $\gamma = 1$ (logarithmic utility). As can be seen, the tail parameter for firm size is essentially unchanged, but the tails for consumption are substantially thicker.

4 Implementation

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 assumptions regarding

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

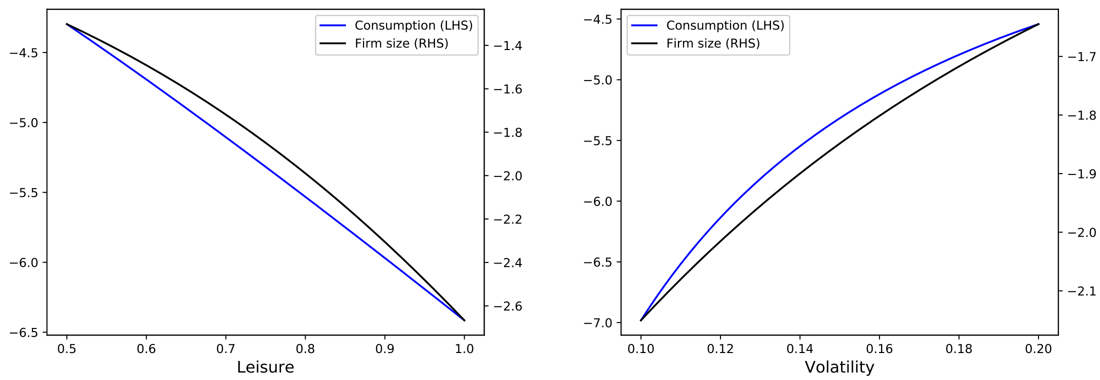

Figure 4: Tail parameters for firm size and consumption with $\overline{\gamma} = 2$.

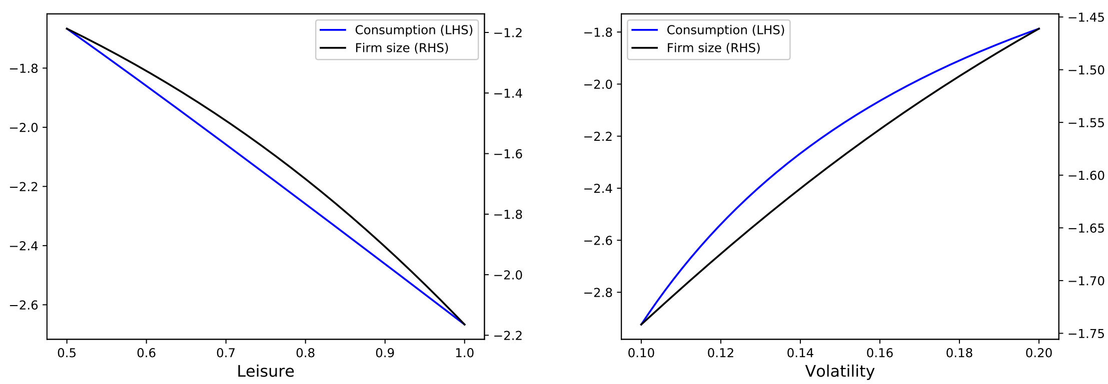

Figure 5: Tail parameters for firm size and consumption with $\overline{\gamma} = 1$.

the contracts agents are capable of writing (and enforcing). Indeed, at one extreme, by extending the results of Atkeson and Lucas (1992), one may show that when agents may write contracts of arbitrary complexity with intermediaries, the optimal policy consists only of lump-sum transfers between agents. Despite the large degree of ex-post heterogeneity, if the government is no more capable of overcoming agency frictions than the private sector, then its role is reduced to addressing the ex-ante differences between agents. However, the assumption of a sector of intermediaries able to commit to contracts for the entirety of an agent's life is unreasonably strong. Not only are these contracts history-dependent and executed over long periods of time, but they also require the intermediary to be capable of observing the asset positions of the agent at all times.

For this reason in this final section I consider a simpler, more realistic market structure and characterize the optimal policy of the government within a parametric class. I assume that agents may save in a risk-free bond and sell shares of their firm with risk-neutral investors (mutual funds), and that the government may levy linear taxes on income and wealth together with lump-sum transfers to workers. 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 primary result of this section is that for this market structure, taxes on profits, (personal) capital income, and wealth play conceptually distinct and complementary roles. The tax on profits achieves redistributive aims, the tax on capital income affects the incentives of owners to retain equity in their firm and hence the degree of risk-sharing in share markets, and the tax on wealth affects the degree of consumption smoothing over time.

The state variable for the agent will consist of firm productivity θ_t as well as financial wealth a_t. The latter is the sum of risk-free savings b_t and firm equity $p_t x_t$, where x_t is the fraction of the firm owned by the agent and p_t the (endogenous) price of the firm,

$$a_t := b_t + p_t x_t. \tag{15}$$

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 intermediaries who trade actuarially fair annuity contracts with agents, promising a return ρ_D on the risk-free savings of agents in exchange for taking possession of these savings at death. Since the mutual fund sector is assumed to be competitive, the price per share will equal the expected discounted value of firm profits, given the funds' expectations regarding the effort of the entrepreneur. The equilibrium notion outlined below will impose the consistency requirement that these beliefs are correct, so that the mutual funds break even in expectation. When the mutual fund believes that the effort exerted is $(\hat{l}_t)_{t\geq 0}$, they value the firm 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 θ_s is productivity at date s conditional on survival, so that the mutual funds effectively discount at rate $r + \rho_D$. If $\hat{l}, w$ and τ_d are constant, the price of the firm is simply

$$p_t = (1 - \tau_d) \frac{\overline{Z}(w)\theta_t}{r + \rho_D - \mu_\theta(\hat{l})} \tag{16}$$

for all $t \geq 0$ almost surely. Note that since I assume that the share market is competitive, investors will make zero expected profits from every trade with entrepreneurs. The following allocation and prescriptions for taxes are therefore unchanged if we instead assume that the share purchases are executed by the government rather than private agents. Under this latter interpretation, government policy serves to partially alleviate frictions in capital markets, just as the taxation of labor income allows government policy to partially substitute for the absence of insurance in the labor market.

When effort expectations $\hat{l}$ and taxes $(\tau_d, \tau_s, \tau_{cg}, \tau_a)$ on dividends, savings, capital gains, and wealth are constant, the entrepreneur's wealth satisfies

$$da_t = [(1 - \tau_s)(r + \rho_D) - \tau_a - \overline{c}_t] a_t dt + \overline{\iota}_t a_t dR_t(l_t; \hat{l}) \tag{17}$$

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

$$dR_t(l_t; \hat{l}) = (\tau_s(r + \rho_D) + (1 - \tau_{cg})\mu_\theta(l_t) - \mu_\theta(\hat{l}))dt + \sigma(1 - \tau_{cg})dZ_t. \tag{18}$$

To understand the law of motion of wealth, note that the expected after-tax excess return on investing in the agent's business is $(1 - \tau_d)\overline{Z}\theta_t/p_t + (1 - \tau_{cg})\mu_\theta(l_t) - (1 - \tau_s)(r + \rho_D)$, which when combined with (16) gives (18).[9] It is important to note the distinction in (17) between $\hat{l}$ and l: the former refers to the level of effort expected by mutual funds, while the latter refers to the true effort exerted by the agent. The two must coincide in equilibrium but only the latter is chosen by the agent. Note also that the tax on firm profits and the wage appear nowhere in the law of motion (17), since these only affect the value of the firm and hence the initial wealth of the agent, and not its evolution over time.

Definition 4.1. Given an interest rate r, linear taxes $\tau = (\tau_d, \tau_s, \tau_{cg}, \tau_a)$, and expectations

[9] 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 C.1.

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

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

Notice that the homotheticity of preferences ensures that the sole effect of a constant linear profits tax is to scale the 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. For simplicity I will suppose that the government borrows and lends to the agents at rate $r = \rho_S$ and runs a primary surplus or deficit to balance its budget.

Definition 4.2. Given constant linear taxes τ, a competitive stock market equilibrium consists of wages w, effort expectations $\hat{l}$, 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 consumer problem in Definition 4.1 given the expectations for leisure $\hat{l}$ and taxes τ.
- The mutual funds 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 simply be set so that the goods market clearing condition is satisfied. By Walras' law, the government's budget constraint is automatically satisfied and so details of the level of debt and transfers are relegated to Appendix C.4. The first observation relevant for the characterization of equilibrium is that if effort expectations are independent of firm productivity, then the latter variable drops out of the individual agent's problem. Conditional on the choice of effort, it becomes a standard portfolio problem of Merton-Samuelson type and so admits a homogeneous solution.

Lemma 4.1. *With linear taxes and constant effort expectations, the choice of leisure is constant, and the policy functions for consumption and investment are of the form $c(a) = \overline{c}a$ and $\iota(a) = \overline{\iota}a$ for some constants $\overline{c}$ and $\overline{\iota}$, where $(\overline{c}, l)$ solve*

$$\begin{aligned}
\frac{\overline{c}}{\rho} &= \frac{(1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s\rho}{E(l)\overline{\gamma}\sigma^2(1-\tau_{cg})} \\
\frac{\rho - \overline{\gamma c}}{1-\overline{\gamma}} &= (1-\tau_s)\rho - \tau_a + \frac{((1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s\rho)^2}{2\overline{\gamma}\sigma^2(1-\tau_{cg})^2}
\end{aligned} \tag{19}$$

and investment is given by

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

The proof of Lemma 4.1 is contained in Appendix C.2. If we impose $\bar{l} = l$, then Lemma 4.1 completely determines the law of motion of the agent's wealth and consumption, given the taxes on income and wealth. Prior to characterizing the optimal linear taxes, we first seek an intuitive understanding of the forces determining consumption and investment in stock market equilibria. 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

$$\overline{\iota} = \frac{\tau_s\rho - \tau_{cg}\mu_\theta(l)}{\overline{\gamma}\sigma^2(1-\tau_{cg})^2}. \tag{20}$$

The expression (20) illustrates some important properties of the agent's problem in the presence of asset-specific taxes. First, the *excess* return on the agent's business is the difference between the taxes paid per unit of savings and the mean return on capital gains. In the absence of such taxes, the expected return on the investment would equal that on the bond, since in this case both the outside investor and the entrepreneur value the firm according to the present discounted value of its after-tax dividends. This illustrates an important point: taxes on various forms of capital income alter the incentives of the entrepreneur to issue shares in her business, by changing both the risk and return on associated capital gains and the return on the alternative investment, the risk-free bond. Albanesi (2006) makes a similar observation in a two-period model with no workers and (physical) capital distributed equally across entrepreneurs, emphasizing that the need to provide incentives for effort provides a justification for the double taxation of capital income.

Note that since the allocations with linear taxes imply constant effort by the agent, given the planner's objective they are necessarily weakly dominated by the restricted-action allocations characterized in Proposition 2.3. Since the policy functions for consumption and investment are linear in wealth, the associated allocations share the property of the restricted-action allocations that they exhibit constant mean and volatility of consumption growth. It follows that if we can find linear taxes such that the mean and volatility of consumption growth in the competitive equilibria coincide with their counterparts in the restricted-action allocations, then we will have found the optimal linear taxes. The fact that the growth in consumption in the restricted-action allocations is characterized by two constants (mean and volatility) implies that in general we have one degree of freedom in our choice of the three tax rates. 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 4.2. *The optimal allocation with linear taxes may be implemented with common taxes on capital income* $\tau_k := \tau_s = \tau_{cg}$ *and wealth* τ_a*, where*

$$\tau_k = \frac{\overline{\gamma}\sigma^2 E(l_{\mathrm{r}}^*)x(l_{\mathrm{r}}^*)}{\rho - \mu_\theta(l_{\mathrm{r}}^*) + \overline{\gamma}\sigma^2 E(l_{\mathrm{r}}^*)x(l_{\mathrm{r}}^*)}$$

$$\tau_a = \overline{\gamma}^2\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2 - \rho\tau_k,$$

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 3.5.

Proof sketch. The mean and volatility of consumption growth in the optimal restricted-action allocation are $(1-\overline{\gamma})\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2/2$ and $\sigma E(l_{\mathrm{r}}^*)x(l_{\mathrm{r}}^*)$, respectively. Lemma 4.1

implies that the equilibrium volatility of consumption is

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

and equating this expression with the efficient analogue gives the tax on capital income. For this tax the policy function in Lemma 4.1 becomes

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

By Lemma 4.1, for this capital tax the drift in wealth (and hence consumption) in the competitive equilibrium allocation is given by

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

Equating this with the drift in consumption in the optimal restricted-action allocation and using the defining equality for $x(l_{\mathrm{r}}^*)$ once more gives

$$\begin{aligned}\tau_a &= -\rho x(l_{\mathrm{r}}^*) - (1-\overline{\gamma})\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2/2 + (1-\tau_k)\rho + \overline{\gamma}\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2 \\ &= ((\overline{\gamma}-1)(\overline{\gamma}-1/2) - (1-\overline{\gamma})/2 + \overline{\gamma})\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2 - \tau_k\rho\end{aligned}$$

which simplifies to the claimed wealth tax. □

The main insight of Proposition 4.2 is that the optimal linear taxation policy calls for separate taxes on wealth and capital income. The debt policy of the government, the level of transfers, and the dividends tax are of secondary importance and so the formal expressions are relegated to Appendix C.4. The redistributive tool of the planner is the tax on firm profits. In contrast, the taxes on capital income and wealth serve instead to ensure that effort and consumption smoothing are at their constrained-efficient levels and are only necessary due to the presence of agency frictions. To see this, note that if agency frictions are absent, either because effort is inelastically supplied $\alpha = 0$ or there is no risk in productivity growth, then $\sigma E(l) \equiv 0$ and the taxes in Proposition 4.2 vanish, as expected. Further, whether or not the 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 4.3. *The revenue raised from the taxes on interest, wealth and capital gains per unit of wealth is* $\overline{\gamma}(\overline{\gamma}-1)\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2$. *Further, if* $\gamma > 1$ *this quantity is increasing in the number of workers per entrepreneur.*

Proof. Using Lemma 4.1 and Proposition 4.2, the taxes raised on interest, wealth, and capital gains are $\tau_a + \rho\tau_k + \tau_k(\mu_\theta(l_{\mathrm{r}}^*) - \rho)\overline{\iota} = \overline{\gamma}^2\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2 - \overline{\gamma}\sigma^2 E(l_{\mathrm{r}}^*)^2 x(l_{\mathrm{r}}^*)^2$, which simplifies to the desired expression. The second claim then follows from Proposition 3.5, 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. $\square$

Corollary 4.3 shows that the personal taxes (i.e., excluding the profit taxes) raise revenue in the aggregate precisely when the expected consumption growth is negative. In particular, in the case of logarithmic utility, the revenue raised from the capital income and wealth taxes nets to zero in the aggregate, which shows that although the tax on capital income is unambiguously positive, the tax on wealth may assume either sign.

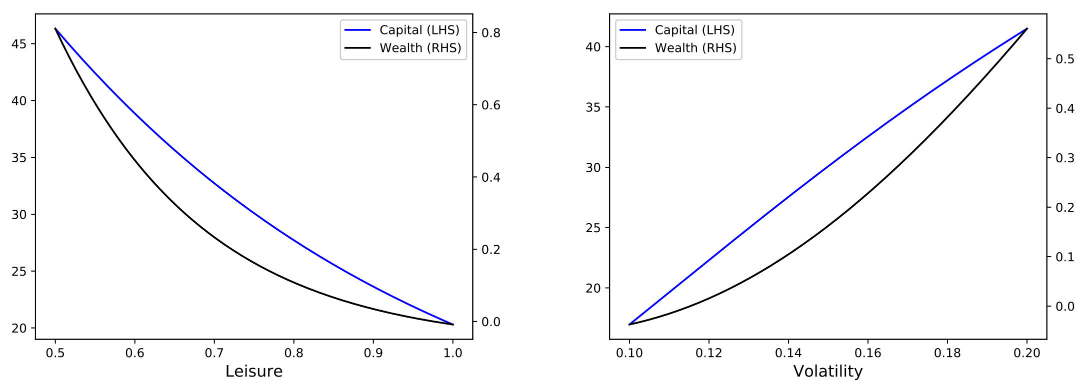

Figure 6: Optimal linear taxes (in percent) with $\overline{\gamma} = 2$.

Numerical illustration. I now compute the optimal linear taxes and revenue raised associated with the parameters adopted in Figure 4. Figure 6 plots the optimal linear taxes in this environment and Figure 7 plots the revenue raised from these taxes, per unit of wealth. Figure 8 repeats this exercise with logarithmic utility.[10] The primary focus of this paper is the theoretical determinants of optimal risk-sharing and taxes when business

[10]The analogue of Figure 7 is omitted since revenue vanishes in this case by Corollary 4.3.

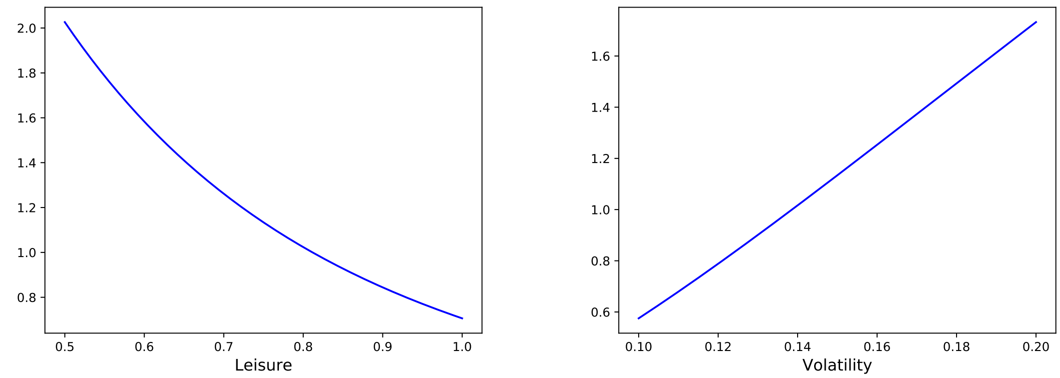

Figure 7: Revenue from optimal linear taxes (as percentage of wealth).

income depends on the history of owner effort, and so the above numerical results do not translate immediately into policy prescriptions. However, it is worth emphasizing that standard parameters over preferences and the volatility of output generate large values for the optimal taxes and the implied revenue raised, *independently* of the level of leisure (which will be sensitive to parameters that are difficult to discipline, such as $\underline{l}$). This suggests that the forces modeled here have a first-order impact on the determination of optimal taxes.

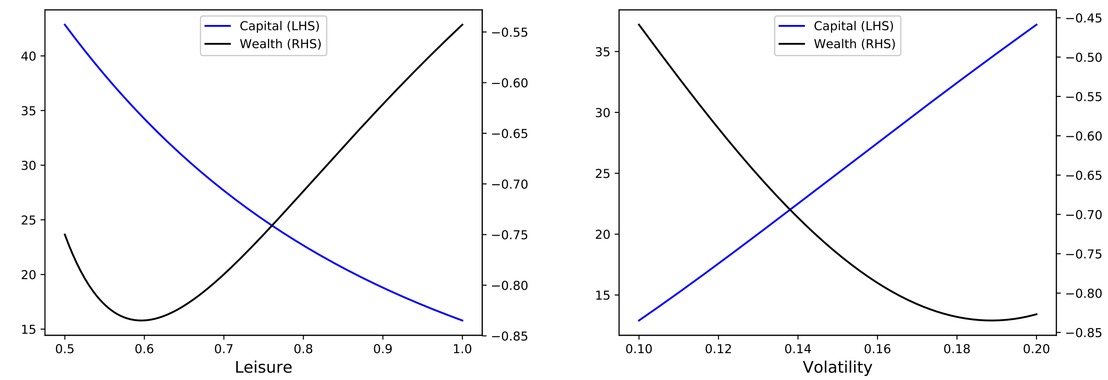

Figure 8: Optimal linear taxes (in percent) with $\overline{\gamma} = 1$.

5 Conclusion

This paper characterizes a class of constrained efficient allocations in an economy with moral hazard and stochastic growth in business income. There were two principal findings. 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 typically holds in the unrestricted optimal contract. In Proposition 3.6 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 when agents may trade shares in their firms in competitive markets or save in a risk-free bond. In this case Proposition 4.2 uncovers a novel role for taxes when productivity depends on unobserved effort: a tax on (personal) capital 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 savings, and wealth, serving three distinct purposes. The tax on profits plays a redistributive role, the tax on risk-free savings provides incentives for retaining equity and continued effort, and the wealth tax serves to implement the efficient level of consumption smoothing.

References

Ai, Hengjie, Dana Kiku, and Rui Li (2016). "A Quantitative Model of Dynamic Moral Hazard." SSRN Scholarly Paper ID 2801116, Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.2801116.

Albanesi, Stefania (2006). "Optimal Taxation of Entrepreneurial Capital with Private Information." Technical Report w12419, National Bureau of Economic Research. doi:10.3386/w12419.

Ales, Laurence, Antonio Andrés Bellofatto, and Jessie Jiaxu Wang (2017). "Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates." Review of Economic Dynamics, 26, pp. 62–90. doi:10.1016/j.red.2017.02.007.

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

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

Best, Michael and Henrik Kleven (2012). "Optimal Income Taxation with Career Effects of Work Effort." SSRN Scholarly Paper ID 2051214, Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.2051214.

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

Chari, V. V. and Patrick J. Kehoe (1999). "Optimal fiscal and monetary policy." In Handbook of Macroeconomics, volume 1, pp. 1671–1745. Elsevier. doi:10.1016/S1574-0048(99)10039-9.

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

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

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

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

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

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

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

Guvenen, Fatih, Gueorgui Kambourov, Burhanettin Kuruscu, Sergio Ocampo-Diaz, and Daphne Chen (2019). "Use It or Lose It: Efficiency Gains from Wealth Taxation." Technical Report w26284, National Bureau of Economic Research. doi:10.3386/w26284.

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

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

Judd, Kenneth L. (1985). "Redistributive taxation in a simple perfect foresight model." Journal of Public Economics, 28(1), pp. 59–83. doi:10.1016/0047-2727(85)90020-9.

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

Kushner, Harold J. and Paul G. Dupuis (2001). Numerical Methods for Stochastic Control Problems in Continuous Time. Springer Science & Business Media.

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

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

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

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

Phelan, Tom (2019). "On the Optimality of Differential Asset Taxation." Working Papers, (WP 19-17). doi:10.26509/frbc-wp-201917.

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

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

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

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

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

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

Stantcheva, Stefanie (2017). "Optimal Taxation and Human Capital Policies over the Life Cycle." Journal of Political Economy, 125(6), pp. 1931–1990. doi:10.1086/694291.

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 $Z^l_t(\omega) := \omega(t) - \sigma^{-1}\int_0^t \mu_\theta(\tilde{l}_s((\omega(s'))_{0\leq s'\leq s}))ds$ for $\omega \in \Omega$, and note that for any l and l' we have $dZ^{l'}_t = dZ^l_t - \sigma^{-1}[\mu_\theta(l'_t) - \mu_\theta(l_t)]$. 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

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

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 $V^{c,l} := (V^{c,l}_t)_{t\geq 0}$ by

$$V^{c,l}_t := \rho\mathbb{E}^l\left[\int_0^\infty e^{-\rho s}u(c_s, l_s)ds \,\middle|\, \mathcal{F}_t\right]. \tag{21}$$

The law of iterated expectations ensures that $V_t^{c,l}$ is a martingale on $(\Omega, \mathcal{F}, P^l)$. The augmented 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 (21) 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 $\mathbb{E}^l \left[\int_0^t e^{-\rho s} S_s dZ_s^l\right] = 0$ for all $t \geq 0$, we have

$$\mathbb{E}^l[V_t^l] = \hat{V}_0^l + \rho \mathbb{E}^l \left[\int_0^t e^{-\rho s} \big(S_s \mu_\theta(l_s) + u(c_s, l_s) - [S_s \mu_\theta(l_s^P) + u(c_s, l_s^P)]\big) ds\right]. \tag{22}$$

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 (22) 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

$$z_t := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}} \qquad\qquad u_t := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}} \theta_t^{-1}. \tag{23}$$

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

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

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

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

$$dU_t = \rho(U_t - u(c_t, l_t)) dt + \sigma(1-\alpha) E(l_t) \big(c_t^{1-\alpha} l_t^\alpha\big)^{1-\gamma} dZ_t. \tag{24}$$

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 (24) becomes $dU_t/U_t = \mu_U dt + \sigma_U dZ_t$ where

$$\mu_U = \rho\big(1 - (\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}\big) \qquad \sigma_U = \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}(1-\overline{\gamma}). \tag{25}$$

We now wish to derive the laws of motion of z_t and u_t in (23). If $f(U_t) := [(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}$ then

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

and so Ito's lemma and (25) imply $dz_t = \mu_z z_t dt + \sigma_z z_t dZ_t$, where

$$\mu_z = \frac{\mu_U}{1-\overline{\gamma}} + \frac{\sigma_U^2}{2}\frac{\overline{\gamma}}{(1-\overline{\gamma})^2} \qquad \sigma_z = \frac{\sigma_U}{1-\overline{\gamma}}. \tag{26}$$

Substituting (25) into (26) gives

$$\begin{aligned} \mu_z &= \rho\left(\frac{1-(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}}{1-\overline{\gamma}}\right) + \frac{\overline{\gamma}\sigma^2}{2}E(l_t)^2(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{2-2\gamma} \\ \sigma_z &= \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}. \end{aligned} \tag{27}$$

Using Ito's lemma for quotients[11] and (27) we have $du_t/u_t = \mu_u dt + \sigma_u dZ_t$ where

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

Factorization and simplification implies that

$$\begin{aligned} \mu_u &= \rho\left(\frac{1-(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}}{1-\overline{\gamma}}\right) + (\overline{\gamma}-1)\frac{\sigma^2}{2}E(l_t)^2(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{2-2\gamma} \\ &\quad + \frac{\sigma_\theta^2(l_t)}{2}(E(l_t)(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma} - 1)^2 - \mu_\theta(l_t) + \frac{\sigma_\theta^2(l_t)}{2} \end{aligned} \tag{28}$$

which in turn implies (9). 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 (23) is $z_t = \exp(U/(1-\alpha))$ and $u_t = \exp(U/(1-\alpha))\theta^{-1}$. Writing $c_t = \overline{c}_t z_t$, we have $U_t - (1-\alpha)\ln c_t = -(1-\alpha)\ln \overline{c}_t$, and so (24) becomes

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

[11]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 $\sigma_X = \sigma E(l_t)(\overline{c}_t^{1-\alpha}l_t^{\alpha})^{1-\gamma}$ and $\sigma_Y = \sigma 1_{l<1}$.

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

$$\begin{aligned} \mu_z &= \frac{\mu_U}{1-\alpha} + \frac{\sigma_U^2}{2(1-\alpha)^2} = -\rho\left(\ln \bar{c}_t + \frac{\alpha}{1-\alpha}\ln l_t\right) + \frac{\sigma^2}{2}E(l_t)^2 \\ \sigma_z &= \frac{\sigma_U}{1-\alpha} = \sigma E(l_t). \end{aligned} \tag{29}$$

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

$$\begin{aligned} \mu_u &= -\rho \ln \bar{c}_t - \frac{\rho\alpha}{1-\alpha}\ln l_t + \frac{\sigma^2}{2}E(l_t)^2 - \mu_\theta(l_t) + \sigma_\theta(l_t)^2 - \sigma^2 E(l_t) \\ &= -\rho \ln \bar{c}_t - \frac{\rho\alpha}{1-\alpha}\ln l_t - \mu_\theta(l_t) + \frac{\sigma_\theta(l_t)^2}{2}(E(l_t)-1)^2 + \frac{\sigma_\theta(l_t)^2}{2}. \end{aligned} \tag{30}$$

A.3 Hamilton-Jacobi-Bellman equation

Using the original choice variable (c, l) and (24), the Hamilton-Jacobi-Bellman equation is given by

$$\begin{aligned} \rho V(U,\theta) = \max_{\substack{c\geq 0 \\ l\in[\underline{l},1]}} \ & \theta 1_{l<1} - c + \rho(U - u(c,l))\frac{\partial V}{\partial U} + \frac{\sigma^2}{2}\Big((1-\alpha)E(l)\big(c^{1-\alpha}l^\alpha\big)^{1-\gamma}\Big)^2\frac{\partial^2 V}{\partial U^2} \\ & + \mu_\theta(l)\theta\frac{\partial V}{\partial \theta} + \frac{\sigma_\theta(l)^2\theta^2}{2}\frac{\partial^2 V}{\partial \theta^2} + \sigma(1-\alpha)E(l)\big(c^{1-\alpha}l^\alpha\big)^{1-\gamma}\sigma_\theta(l)\theta\frac{\partial^2 V}{\partial U \partial \theta}. \end{aligned} \tag{31}$$

If $\gamma > 1$, then in terms of the variables $(\bar{c}, l)$, where $c = \bar{c}[(1-\gamma)U_t]^{\frac{1}{1-\overline{\gamma}}}$, (31) becomes

$$\begin{aligned} \rho V(U,\theta) = \max_{\substack{\bar{c}\geq 0 \\ l\in[\underline{l},1]}} \ & \theta 1_{l<1} - \bar{c}[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}} + \rho\Big(1 - \big(\bar{c}^{1-\alpha}l^\alpha\big)^{1-\gamma}\Big)U\frac{\partial V}{\partial U} \\ & + \frac{\sigma^2}{2}\Big((1-\overline{\gamma})E(l)\big(\bar{c}^{1-\alpha}l^\alpha\big)^{1-\gamma}\Big)^2 U^2\frac{\partial^2 V}{\partial U^2} + \mu_\theta(l)\theta\frac{\partial V}{\partial \theta} \\ & + \frac{\sigma_\theta(l)^2\theta^2}{2}\frac{\partial^2 V}{\partial \theta^2} + \sigma(1-\overline{\gamma})E(l)\big(\bar{c}^{1-\alpha}l^\alpha\big)^{1-\gamma}U\sigma_\theta(l)\theta\frac{\partial^2 V}{\partial U\partial\theta} \end{aligned} \tag{32}$$

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 $c = \bar{c}\exp(U/(1-\alpha))$, (32) becomes

$$\begin{aligned} \rho V(U,\theta) = \max_{\substack{\bar{c}\geq 0 \\ l\in[\underline{l},1]}} \ & \theta 1_{l<1} - \bar{c}\exp(U/(1-\alpha)) + \rho(-(1-\alpha)\ln\bar{c} - \alpha\ln l)\frac{\partial V}{\partial U} \\ & + \frac{\sigma^2}{2}(1-\alpha)^2E(l)^2\frac{\partial^2 V}{\partial U^2} + \mu_\theta(l)\theta\frac{\partial V}{\partial\theta} + \frac{\sigma_\theta(l)^2\theta^2}{2}\frac{\partial^2 V}{\partial\theta^2} + \sigma(1-\alpha)E(l)\sigma_\theta(l)\theta\frac{\partial^2 V}{\partial U\partial\theta}. \end{aligned} \tag{33}$$

Proposition A.1. *When $\gamma > 1$ the solution to the Hamilton-Jacobi-Bellman equation (32) is of the form $V(U,\theta) = v(u)\theta$ for some function v solving*

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

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

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

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

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

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

and

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

Substitution into (32) gives

$$\begin{aligned}
\rho v(u)\theta = \max_{\substack{\overline{c}\geq 0\\ l\in[\underline{l},1]}} \; & (1_{l<1} - \overline{c}u)\theta + \rho\Big(1 - \big(\overline{c}^{1-\alpha}l^{\alpha}\big)^{1-\gamma}\Big)U\frac{\theta^{\overline{\gamma}}u^{\overline{\gamma}}}{1-\alpha}v'(u)\\
&+ \frac{\sigma^2}{2}E(l)^2(1-\overline{\gamma})^2\big(\overline{c}^{1-\alpha}l^{\alpha}\big)^{2-2\gamma}U^2\frac{(u\theta)^{2\overline{\gamma}-2}}{(1-\alpha)^2}[\overline{\gamma}uv'(u) + u^2v''(u)]\theta\\
&+ \mu_\theta(l)\theta(v(u) - uv'(u)) + \frac{\sigma_\theta(l)^2\theta}{2}v''(u)u^2 - \sigma\sigma_\theta(l)E(l)\big(\overline{c}^{1-\alpha}l^{\alpha}\big)^{1-\gamma}v''(u)u^{1+\overline{\gamma}}(1-\gamma)U\theta^{\overline{\gamma}}.
\end{aligned}$$

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

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

which simplifies as claimed upon factorization. □

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* (33) *is of the form $V(U,\theta) = v(u)\theta$ for some function v solving*

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

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

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

Combining this with the associated first-order condition gives the pair of equations

$$\begin{aligned}\rho\overline{v}_{\mathrm{r}}(l) &= -\overline{c} + \rho\bigg(\frac{1-(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\bigg)\overline{v}_{\mathrm{r}}(l) + \frac{\overline{\gamma}\sigma^2}{2}E(l)^2\big(\overline{c}^{1-\alpha}l^\alpha\big)^{2-2\gamma}\overline{v}_{\mathrm{r}}(l)\\ 0 &= -\overline{c} - \rho(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}\overline{v}_{\mathrm{r}}(l) + (2-2\overline{\gamma})\frac{\overline{\gamma}\sigma^2}{2}E(l)^2\big(\overline{c}^{1-\alpha}l^\alpha\big)^{2-2\gamma}\overline{v}_{\mathrm{r}}(l).\end{aligned}$$

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

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

which simplifies to the claimed quadratic. Using the fact that $c_t = \overline{c}u_t\theta_t = \overline{c}z_t$ for z_t is defined in Appendix A.2, the expression (27) and the defining quadratic for x then give

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

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

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

Since $\overline{\gamma}, \gamma > 1$, T is increasing in l whenever $\overline{v}$ 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 $\overline{v}$, 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*

$$\begin{aligned} v_{\mathrm{r}}(u;l) &= \frac{1_{l<1}}{\rho - \mu_\theta(l)} - \frac{1}{\rho}\exp\big(\sigma^2E(l)^2/[2\rho]\big)l^{-\frac{\alpha}{1-\alpha}}u\\ c_{\mathrm{r}}(u;l) &= \exp\big(\sigma^2E(l)^2/[2\rho]\big)l^{-\frac{\alpha}{1-\alpha}}u.\end{aligned} \tag{34}$$

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

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

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,

$$
\begin{aligned}
C_t &:= e^{-\rho_D t}\underline{C}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} C_t^T dT \\
\underline{C}_t &:= \int_{\Omega'} \mathbb{E}[c_t^{v,\theta,i}]\Phi(d\omega), \qquad C_t^T := \sum_{i=E,W} \eta_i \mathbb{E}[c_t^{T,i}] \\
Y_t &:= e^{-\rho_D t}\underline{Y}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} Y_t^T dT \\
\underline{Y}_t &:= \int_{\Omega'} \mathbb{E}[F(\theta_t^{v,\theta,i}, L_t^{v,\theta,i})]\Phi(d\omega), \qquad Y_t^T := \eta_E \mathbb{E}[F(\theta_t^{T,E}, L_t^{T,E})] \\
L_t &:= e^{-\rho_D t}\underline{L}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} L_t^T dT \\
\underline{L}_t &:= \int_{\Omega} \mathbb{E}[L_t^{v,\theta,E}]\Phi(d\omega), \qquad L_t^T = \eta_E \mathbb{E}[L_t^{T,E}].
\end{aligned}
$$

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

$$
\begin{aligned}
U_t &:= e^{-\rho_D t}\underline{U}_t + \rho_D \int_0^t e^{-\rho_D(t-T)} U_t^T dT \\
\underline{U}_t &:= \int_{\Omega'} \mathbb{E}[u(c_t^{v,\theta,i}, l_t^{v,\theta,i})]\Phi(d\omega), \qquad U_t^T := (1-\eta_E)\mathbb{E}[u(c_t^{T,W}, l_t^{T,W})]
\end{aligned}
$$

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

$$
\begin{aligned}
U^P &= \int_{\Omega'} \mathbb{E}\left[\int_0^\infty e^{-\rho t} u(c_t^{v,\theta,i}, l_t^{v,\theta,i}) dt\right]\Phi(d\omega) \\
&\quad + \rho_D \int_0^\infty e^{-\rho_S T}(1-\eta_E)\mathbb{E}\left[\int_T^\infty e^{-\rho(t-T)} u(c_t^{T,W}, l_t^{T,W}) dt\right] dT \\
&= \int_0^\infty e^{-\rho_S t} \int_{\Omega'} e^{-\rho_D t}\mathbb{E}[u(c_t^{v,\theta,i}, l_t^{v,\theta,i})] dt \Phi(d\omega) dt \\
&\quad + \rho_D(1-\eta_E) \int_0^\infty e^{-\rho_S t} \int_0^t e^{-\rho_D(t-T)}\mathbb{E}[u(c_t^{T,W}, l_t^{T,W})] dT dt
\end{aligned}
$$

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. $\square$

B.2 Reduction to principal-agent problem

The Lagrangian associated with the relaxed planner's problem is

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

Using the above expressions we may then expand this as

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

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 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 is forced to provide 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

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

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 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.[12]

Proof of Lemma 3.2. The process $(\theta_t, u_t)_{t\geq 0}$ is a diffusion process driven by the same Brownian

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

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

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

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 (11). □

B.4 Restricted-action distributions

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

Lemma B.2. *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*

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

where $\alpha_X^{\pm} = \overline{\mu}_X/\sigma_X^2 \pm \sqrt{(\overline{\mu}_X/\sigma_X)^2 + 2\delta_X/\sigma_X^2}$*, 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 3.6. The proof proceeds by combining Lemma B.2 with the policy functions for consumption given in Proposition 2.3. From (27) we have $dc_t = \mu_c c_t dt + \sigma_c c_t dZ_t$, where

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

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

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

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

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

which simplifies as claimed using the defining quadratic for x. □

C Implementation

C.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,

$$\begin{aligned}
a_{t+\Delta} &= \text{savings} + \text{interest} - \text{consumption} + \text{dividends} + \text{stocks} + \text{capital gains} \\
&= b_t + (1-\tau_s)(r+\rho_D)\Delta b_t - \Delta c_t + \Delta(1-\tau_d)\overline{Z}\theta_t x_t + p_t x_t + (1-\tau_{cg})(p_{t+\Delta} - p_t)x_t \\
&= a_t + \Delta\big[(1-\tau_s)(r+\rho_D)a_t - c_t + [(1-\tau_d)\overline{Z}\theta_t/p_t - (1-\tau_s)(r+\rho_D)]\iota_t\big] \\
&+ (p_{t+\Delta}/p_t - 1)(1-\tau_{cg})\iota_t.
\end{aligned} \tag{35}$$

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 (35) and taking limits as $\Delta \to 0$

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

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

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

which is exactly the expression implied by (17) and (18).

C.2 Agent problems

Proof of Lemma 4.1. When the agent faces constant taxes and expectations of constant effort $\hat{l}$, their value function solves the Hamilton-Jacobi-Bellman equation

$$\begin{aligned}
\rho V(a) = \max_{\overline{c},l,\overline{\iota}} \rho\frac{((\overline{c}a)^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\gamma} &+ [-\tau_a - \overline{c} + (1-\tau_s)(r+\rho_D)]aV'(a) \\
&+ \Big[(1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r+\rho_D)\Big]\overline{\iota}aV'(a) + \frac{\sigma^2a^2}{2}\overline{\iota}^2(1-\tau_{cg})^2V''(a).
\end{aligned} \tag{36}$$

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

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

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

$$\begin{aligned}
(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma} &= \frac{\overline{c}}{\rho}\overline{V}(\hat{l}) \\
(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma} &= \frac{l}{\rho\alpha}(\overline{\mu}_0 - \overline{\mu}_1)(1-\alpha)(1-\tau_{cg})\overline{\iota}\overline{V}(\hat{l}) = \frac{1}{E(l)}(1-\tau_{cg})\overline{\iota}\overline{V}(\hat{l}) \\
\overline{\iota} &= \frac{(1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r+\rho_D)}{\overline{\gamma}\sigma^2(1-\tau_{cg})^2}.
\end{aligned}$$

Substituting the third into the second gives

$$\frac{\overline{c}}{\rho} = \frac{(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\overline{V}(\hat{l})} = \frac{(1-\tau_{cg})\mu_\theta(l) - \mu_\theta(\hat{l}) + \tau_s(r+\rho_D)}{E(l)\overline{\gamma}\sigma^2(1-\tau_{cg})}. \tag{37}$$

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

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

We then have $\overline{V}(\hat{l}) = \rho(\overline{c}^{1-\alpha}l^{\alpha})^{1-\gamma}/\overline{c}$, where $\overline{c}$ and l solve (37) and

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

as desired. □

C.3 Efficient policy functions and utility

In this section we determine the policy functions and value functions when taxes on capital income and wealth are given by Proposition 4.2.

Lemma C.1. *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* $\bar{\iota} = \big(1 + \overline{\gamma}\sigma^2 E(l)x(l)/(\rho - \mu_\theta(l))\big)E(l)x(l)$*.*

Proof. If $l = \hat{l}$ and $\tau_s = \tau_{cg} = \tau_k$ is given in Proposition 4.2 then (19) becomes

$$\frac{\overline{c}}{\rho} = \frac{\tau_k(\rho - \mu_\theta(l_{\mathrm{r}}^*))}{E(l_{\mathrm{r}}^*)\overline{\gamma}\sigma^2(1-\tau_k)} = \frac{\sigma_a}{E(l_{\mathrm{r}}^*)\sigma} = x(l_{\mathrm{r}}^*) \tag{38}$$

which gives the consumption function, and the expression for utility then follows from this and $\overline{V} = (\overline{c}^{1-\alpha}l^{\alpha})^{1-\gamma}[\rho/\overline{c}]$. Finally, using Proposition 4.2 once more implies

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

Substituting into (20) then gives

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

which simplifies as claimed. □

C.4 Equilibrium characterization

To complete the characterization described in Proposition 4.2, in this section I specify the transfers to workers, tax on dividends and level of outstanding government debt necessary to implement the optimal restricted-action allocation. Lemma C.1 shows that for the taxes given in Proposition 4.2 utility is $\overline{V}^{\frac{1}{1-\overline{\gamma}}}a = \rho x(l)^{-\frac{\overline{\gamma}}{1-\overline{\gamma}}} l^{\frac{\alpha}{1-\alpha}} a$. For the market-clearing wage we have

$$\overline{Z} = (1-\beta)Z\biggl(\frac{\rho_D\eta_E}{\rho_D-\mu_\theta(l)}\biggr)^{-\beta}\overline{L}^\beta = \frac{(\rho_D-\mu_\theta(l))}{\eta_E\rho_D}(1-\beta)Y. \tag{39}$$

The utility that agents obtain is then a solution to

$$Y = Z\biggl(\frac{\rho_D\eta_E}{\rho_D-\mu_\theta(l)}\biggr)^{1-\beta}\overline{L}^\beta = \biggl(\frac{\rho_D\eta_E\overline{c}_{\mathrm{r}}(l)}{\rho_D-\mu_c(l)}+1-\eta_E\biggr)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}} =: \kappa(l)[(1-\gamma)U]^{\frac{1}{1-\overline{\gamma}}}.$$

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

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

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

$$\kappa(l) = \frac{\rho_D\eta_E x(l)^{\frac{1}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}}{\rho_D-(1-\overline{\gamma})\sigma^2E(l)^2x(l)^2/2}+1-\eta_E. \tag{40}$$

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 $(Y/\rho)\Bigl(x(l)^{\frac{\overline{\gamma}}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}/\kappa(l)-\beta\Bigr)$. The value of the firm is $(1-\tau_d)\overline{Z}/(\rho-\mu_\theta(l))$, and so using (39) the tax on dividends is

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

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

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

$$\tau_d(1-\beta)-\rho_D(1-\eta_E)\frac{T}{Y} = 1-\beta-\frac{\rho_D(\rho-\mu_\theta(l))}{\rho(\rho_D-\mu_\theta(l))}\biggl(\frac{x(l)^{\frac{\overline{\gamma}}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}}{\kappa(l)}-\beta\biggr)\eta_E - \frac{\rho_D}{\rho}(1-\eta_E)\biggl(\frac{1}{\kappa(l)}-\beta\biggr).$$

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

$$1-\frac{\rho_D}{\rho}\Biggl[\frac{(\rho-\mu_\theta(l))}{(\rho_D-\mu_\theta(l))}\biggl(\frac{x(l)^{\frac{\overline{\gamma}}{1-\overline{\gamma}}}l^{-\frac{\alpha}{1-\alpha}}-\kappa(l)}{\kappa(l)(1-\beta)}+1\biggr)\eta_E+(1-\eta_E)\biggl(\frac{1/\kappa(l)-\beta}{1-\beta}\biggr)\Biggr]. \tag{41}$$

It is instructive to examine the above when there are no agency frictions or disutility of labor. All workers are given $Y(1-\beta)/\rho$, the discounted value of profits, and the tax on dividends is

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

The debt held by the government (i.e. the negative of the amount issued) as a fraction of profits is

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

In this case the value of each firm at birth is $\eta_E\overline{Z}/\rho=(1-\beta)Y/\rho$, which coincides with the transfers to workers, and ensures that all agents have the same wealth. To understand who holds the debt in this economy, note that the agents and the government hold $(1-\beta)(Y/\rho)\rho_D\eta_E/(\rho_D-\mu_\theta(l))$ and so the mutual funds must hold the negative of this quantity. Since the mutual funds own all of the firms in the economy, they earn all (after-tax) profits, and so after purchasing new firms with a flow of $\rho_D\eta_E(1-\beta)Y/\rho$ government bonds their net revenues every instant are

$$\begin{aligned}&(1-\beta)(1-\tau_d)Y-\frac{\rho_D}{\rho}\eta_E(1-\beta)Y-\frac{Y}{\rho}\frac{\rho_D\eta_E}{(\rho_D-\mu_\theta(l))}(1-\beta)(\rho-\rho_D)\\&=(1-\beta)\frac{Y}{\rho}\left[\rho-\rho_D+(\rho-\rho_D)\frac{\mu_\theta(l)}{(\rho_D-\mu_\theta(l))}-\frac{\rho_D}{(\rho_D-\mu_\theta(l))}(\rho-\rho_D)\right]\eta_E=0\end{aligned}$$

as expected. It is also instructive to determine the value of the dividend tax in the situation in which $\beta=0$ and there are no workers, to understand the role played by the dividends tax in the decentralization. In this case we have

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

In particular, in the absence of any discounting across generations we have $\rho_D=\rho$, and hence $\tau_d=\mu_c/\rho$, and so by Corollary 4.3 the dividends tax takes the opposing sign of the taxes raised on capital and wealth, and vanishes as agency frictions vanish.

D Numerical method

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 (28) 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

$$\begin{aligned}\mu_u&=\rho\left(\frac{1-(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\overline{\gamma}}\right)+(\overline{\gamma}-1)\frac{\sigma^2}{2}E(l)^2(\overline{c}^{1-\alpha}l^\alpha)^{2-2\gamma}\\&\quad+\frac{1}{2}(\sigma E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}-\sigma_\theta(l))^2-\mu_\theta(l)+\frac{\sigma_\theta^2(l)}{2}\\\sigma_u&=\sigma E(l)(\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}-\sigma_\theta(l)\end{aligned}\tag{42}$$

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

$$\rho\overline{V} = \max_{\substack{\overline{c}\geq 0 \\ l\in[\underline{l},1]}} \theta 1_{l<1} - \overline{c}u\theta + \mu_u u\overline{V}_1 + \frac{\sigma_u^2 u^2}{2}\overline{V}_{11} + \mu_\theta\theta\overline{V}_2 + \frac{\sigma_\theta^2\theta^2}{2}\overline{V}_{22} + \sigma_u u\sigma_\theta\theta\overline{V}_{12}. \tag{43}$$

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

$$\rho\overline{V} = \max_{\substack{\overline{c}\geq 0 \\ l\in[\underline{l},1]}} \theta 1_{l<1} - \overline{c}u\theta + (\mu_u + \sigma_u\sigma_\theta)u\overline{V}_1 + \frac{\sigma_u^2 u^2}{2}\overline{V}_{11} + \mu_\theta\theta\overline{V}_2 + \frac{\sigma_\theta^2\theta^2}{2}\overline{V}_{22}. \tag{44}$$

Solving (44) is more convenient than solving (43) because there are no cross partial derivatives that complicate the construction of approximating chains. The solution to (44) 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

$$(du_t, d\theta_t) = \Big((\mu_u + \sigma_u\sigma_\theta)u_t dt + \sigma_u u_t dZ_t^{(1)}, \mu_\theta\theta_t dt + \sigma_\theta\theta_t dZ_t^{(2)}\Big) \tag{45}$$

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

$$\begin{aligned} p(u+\Delta_u,\theta) &= \frac{\Delta_t}{\Delta_u^2}\left(\frac{\sigma_u^2 u^2}{2} + \Delta_u\hat{\mu}_1 u\right) \\ p(u-\Delta_u,\theta) &= \frac{\Delta_t}{\Delta_u^2}\left(\frac{\sigma_u^2 u^2}{2} + \Delta_u\hat{\mu}_2 u\right) \\ p(u,(1\pm\Delta_\theta)\theta) &= \frac{\Delta_t}{\Delta_\theta^2}\left(\frac{\sigma_\theta^2}{2} + \Delta_\theta\mu_\theta^{\pm}\right) \end{aligned} \tag{46}$$

define a locally consistent Markov chain if Δ_t is sufficiently small. Writing $x(\overline{c},l) = (\overline{c}^{1-\alpha}l^\alpha)^{1-\gamma}$ for brevity, the expressions in (42) imply

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

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

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

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

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

Substituting the above expressions, dividing by Δ_t and simplifying gives

$$\begin{aligned}
0 = &-\frac{1}{\Delta_t}(1-e^{-\rho\Delta_t})v(u) + \max_{\bar{c}\geq 0, l\in[\underline{l},1]} 1_{l<1} - \bar{c}u + e^{-\rho\Delta_t}\mu_\theta v(u) \\
&+ e^{-\rho\Delta_t}\left(\hat{\mu}_1 u v^F - \hat{\mu}_2 u v^B + \sigma_u^2 u^2 v^{C2}/2\right)
\end{aligned} \tag{47}$$

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 (47) 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

$$\begin{aligned}
v(u+\Delta_u) &: \frac{\hat{\mu}_1 u}{\Delta_u} + \frac{\sigma_u^2 u^2}{2\Delta_u^2} \\
v(u) &: -\rho + \mu_\theta - \frac{1}{\Delta_u}(\hat{\mu}_1 + \hat{\mu}_2)u - \frac{\sigma_u^2 u^2}{\Delta_u^2} \\
v(u-\Delta_u) &: \frac{\hat{\mu}_2 u}{\Delta_u} + \frac{\sigma_u^2 u^2}{2\Delta_u^2}
\end{aligned} \tag{48}$$

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

$$\begin{aligned}
\mu_u &= -\rho \ln \bar{c}_t - \frac{\rho\alpha}{1-\alpha}\ln l_t - \mu_\theta(l_t) + \frac{\sigma^2}{2}(E(l_t) - 1_{l_t<1})^2 + \frac{\sigma^2}{2}1_{l_t<1} \\
\sigma_u &= \sigma(E(l_t) - 1_{l_t<1})
\end{aligned} \tag{49}$$

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

$$(du_t, d\theta_t) = \left((\mu_u + \sigma_u\sigma_\theta)u_t dt + \sigma_u u_t dZ_t^{(1)}, \mu_\theta\theta_t dt + \sigma_\theta\theta_t dZ_t^{(2)}\right) \tag{50}$$

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 (46), 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 (42) 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

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

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 (48) 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

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