Risk Markups

Sebastian Di Tella — Stanford GSB & NBER
Cedomir Malgieri — Stanford
Christopher Tonetti — Stanford GSB & NBER

April 4, 2025

Version 0.99

We study optimal policy when heterogeneous markups reflect compensation for uninsurable persistent idiosyncratic risk. Entrepreneurs hire labor trading off expected profits against risk. The optimal labor tax has a keep rate equal to the product of (1) the aggregate markup and (2) workers' consumption share divided by their Pareto weight. The markup component reflects inefficient risk premia and calls for a subsidy, while the consumption-share component reflects inefficient precautionary saving and calls for a tax. In the long run, precautionary saving dominates, generating an inefficient excess of entrepreneur wealth. The optimal tax prevents impoverished workers from supplying too much labor.

1 Introduction

Markups are measured to be large and heterogeneous across firms, which can lower TFP, distort aggregate employment, and create inequality. The role of policy depends on the source of markups. A common view is that markups arise because firms have market power. From this perspective, Baqaee and Farhi (2020) and Edmond, Midrigan, and Xu (2023) show that it is optimal to increase output with a subsidy to undo the markup distortions.

Our main contribution is to study optimal policy when measured markups are instead risk premia. Since at least Knight (1921) economists have recognized that profits may reflect compensation for risk. Providing quantitative support for this view, Boar, Gorea, and Midrigan (2023) attribute the vast majority of the dispersion in firm profits to uninsurable idiosyncratic risk, with some contribution from market power or decreasing returns to scale.

The constrained-efficient allocation implements the first-best level of employment using a labor tax with a keep rate equal to the product of (1) the aggregate markup and (2) workers' consumption share divided by their Pareto weight. The markup component reflects inefficient risk premia and calls for a labor subsidy to eliminate the aggregate markup. This is the same as in the market-power case, even though markups correctly capture the private cost of risk.

But an important insight is that risk markups also distort the wealth distribution relative to the first best. Even though the planner can use lump-sum transfers to achieve any desired redistribution, precautionary saving leads entrepreneurs to overaccumulate wealth over time, leading impoverished workers to supply too much labor. The consumption-share component captures this effect and calls for a labor tax. In the long run, the second effect dominates and the optimal policy is a tax to reduce output, in sharp contrast to the common wisdom derived from the market-power perspective.

Model Overview and Main Results. There is a representative worker and a continuum of entrepreneurs. The workers supply labor, which is the only factor of production, and face no risk. Each entrepreneur is endowed with a distinct linear technology to produce the homogeneous consumption good. There are four essential ingredients. First, entrepreneurs persistently differ in the productivity of their technology, which introduces the possibility of misallocation. Second, hiring labor is risky. Entrepreneurs must hire labor before the realization of their productivity shock. Third, entrepreneurs are risk averse. Fourth, there is incomplete idiosyncratic risk sharing. There are competitive markets for goods, labor, a risk free bond, and firm equity. Motivated by moral hazard arguments, the entrepreneur can sell at most a fraction $1 - \phi$ of the firm value and must retain a ϕ share of the firm equity.

These ingredients generate heterogeneous risk markups. All firms charge the same price for the homogeneous good they sell and pay the same wage to hire labor. However, there is a wedge in the first order condition for hiring labor, such that the price of the good is greater than the unit labor cost. This risk markup is compensation for risk. Hiring labor is like investing in a risky asset and entrepreneurs consider the trade-off between the expected return and the risk associated with hiring.

Even though production of the homogeneous good is linear, the most productive entrepreneur does not hire all the labor because that would expose him to too much risk. There is a sense in which there are no markups in the competitive equilibrium, since the risk-adjusted marginal cost is equal to the price of the final good for all firms. Nonetheless, risk markups generate lower aggregate TFP (misallocation), an aggregate labor wedge, and income inequality among entrepreneurs and between workers and entrepreneurs.

We study the constrained-efficient allocation of a planner who has access to a time-varying tax on labor and lump-sum transfers. The lump-sum transfers allow the planner to achieve any redistribution, allowing us to focus on whether there are efficiency reasons to use the labor tax. We first leverage a separation result to show that the constrained-efficient planner implements the first-best level of employment. To do so, he must use a tax with a keep rate (one minus the tax rate) equal to the product of (1) the aggregate markup and (2) workers' consumption share divided by their Pareto weight.

A key insight of this paper is that the single feature of uninsurable idiosyncratic risk simultaneously generates both inefficient risk markups and inefficient precautionary saving. The markup component of the optimal tax reflects inefficient risk premia and the consumption-share component reflects inefficient precautionary saving.

Even though the risk markup represents compensation for utility-relevant risk taking, it is still inefficient. Individually, as an entrepreneur hires more labor, he is exposed to more risk and consequently has more volatility in his consumption. The planner internalizes that the extra output produced helps all entrepreneurs smooth their consumption. The risk markup correctly reflects the private marginal cost of risk, but not the social cost. This inefficiency calls for the planner to use a labor subsidy to eliminate the aggregate markup, just as in the case where markups reflect market power.

However, inefficient risk premia are not the planner's sole concern. Uninsurable idiosyncratic risk leads to inefficient precautionary saving. Over time, entrepreneurs accumulate too much wealth relative to the first best. The planner takes Euler equations as a constraint, so entrepreneurs must receive a larger share of consumption than is warranted by their Pareto weights. Workers' consumption share is too small and their labor supply is therefore too large relative to the constrained-efficient level of employment (which coincides with the first best). This inefficiency calls for the planner to use a labor tax to prevent workers from supplying too much labor.

In the long run, the precautionary saving effect dominates the risk premia effect and the optimal policy is a tax, in sharp contrast with the market-power perspective. In our simple numerical exercise, the welfare loss from implementing a labor subsidy equal to the aggregate markup is eight times larger than the welfare gain from the optimal policy. More broadly, the risk perspective on measured markups focuses attention on precautionary saving and the inefficiency of the wealth distribution.

1.1 Relationship to Existing Literature

Measured markups are large and heterogeneous across firms and can lower TFP (misallocation).[1] We contribute to the recent literature on risk and misallocation. In David, Schmid, and Zeke (2022), dispersion in the marginal product of capital is generated by heterogeneous risk premia. There is an aggregate shock and different firms have different exogenous loadings on the aggregate risk factor. As in CAPM, the different risk profiles generate different risk premia, which generate deviations from equalizing MPKs, even though the decentralized equilibrium is efficient. Our paper also explores the relationship between risk and misallocation, but in our model risk premia are endogenous, generated by idiosyncratic risk and incomplete markets. This leads to a substantively different perspective on optimal policy. David and Zeke (2024) further explore the implications of aggregate risk and misallocation for optimal monetary policy in a New Keynesian model and David, Ranciere, and Zeke (2023) explore the implications for international diversification and the labor share.

Boar, Gorea, and Midrigan (2023) is a closely related paper that also studies an environment with uninsurable idiosyncratic risk. Theirs is a quantitative paper that explains the source of heterogeneity in firm returns. Their model additionally features decreasing returns to scale production technology (equivalently monopolistic competition) and collateral constraints. They find that uninsurable idiosyncratic risk is by far the most important feature that explains the empirical variation in returns. Motivated by their quantitative results on the importance of uninsurable idiosyncratic risk, we compliment their paper by providing a theoretical analysis of optimal policy in such a world. Since optimal policy is relatively well understood when misallocation is generated by market power or by collateral constraints, we only model uninsurable idiosyncratic risk. This allows us to provide insights from analytical results in a tractable model.

Our model of entrepreneurship is similar to that in Moll (2014) and our analysis of

[1]See Hall (1988), De Loecker and Warzynski (2012), Restuccia and Rogerson (2008), and Hsieh and Klenow (2009).

optimal policy is similar to Itskhoki and Moll (2019). Those papers, however, focus on collateral constraints that prevent firms from operating at their optimal scale, which induces capital misallocation. A key insight from their analysis is that the degree of persistence in the stochastic idiosyncratic productivity process affects the magnitude of steady-state TFP losses from misallocation and the speed of transition towards steady state. Buera and Shin (2013) similarly focus on the effects of collateral constraints on misallocation, in a model with decreasing returns to scale production and an occupational choice, with a focus on quantitative results. To clearly distinguish the new mechanism of risky production in our paper, we make labor the only factor of production and omit collateral constraints.

Our paper is also related to Meh and Quadrini (2006) and Angeletos (2007) in that these papers study the effect of risky production, and not collateral constraints, on long-run allocations. Meh and Quadrini (2006) and Angeletos (2007), however, study a model with iid idiosyncratic risk and focus on capital accumulation. Di Tella and Hall (2022) study labor risk premiums and investment dynamics in the context of business cycles. These models with iid shocks do not feature heterogeneous types and, thus, no potential for misallocation across types. We study a model without aggregate risk, capital, or capital constraints to focus on the long-run relationship between idiosyncratic risk, markups, and misallocation across persistently heterogeneous entrepreneurs.

Eeckhout and Veldkamp (2023) is another paper exploring how risk affects our interpretation of measured markups. In their paper, firms acquiring data can reduce risk markups by reducing risk (improving forecasting) but can increase markups by inducing investment to capture market share and exploit market power. Dou, Ji, Tian, and Wang (2025) also studies how risk and incomplete markets affects misallocation, with a focus on how misallocation introduces medium-run fluctuations in TFP that affect aggregate growth and asset prices. These are all in contrast to papers on misallocation without an essential role for risk, such as those that generate misallocation with market power as in Baqaee and Farhi (2020) and Edmond, Midrigan, and Xu (2023), with search frictions as in Menzio (2024), or adjustment costs as in Asker, Collard-Wexler, and De Loecker (2014).

2 The Model

2.1 The Environment

Time is continuous and infinite, indexed by $t \geq 0$. There is a representative worker and a continuum of entrepreneurs, indexed by $i \in [0,1]$, with preferences

$$U_w(c_w, \ell) = \mathbb{E}\left[\int_0^\infty e^{-\rho_w t}\left(\log c_{wt} - \frac{\ell_t^{1+1/\psi}}{1+1/\psi}\right) dt\right], \tag{1}$$

$$U_e(c_i) = \mathbb{E}\left[\int_0^\infty e^{-\rho_e t} \log c_{it} dt\right]. \tag{2}$$

We focus on the case with entrepreneurs more impatient than workers, $\rho_e > \rho_w > 0$, to obtain a non-degenerate stationary wealth distribution.[2]

Each entrepreneur has a linear technology that uses labor to produce the final good:

$$dY_{it} = \ell_{it}\ (z_{it}dt + \sigma_y dB_{yit}), \tag{3}$$

$$dz_{it} = \mu_z(z_{it})dt + \sigma_z(z_{it})dB_{zit}, \tag{4}$$

where B_{yit} and B_{zit} are two Brownian motions specific to the entrepreneur that have correlation λ_{yz}, and $\sigma_y, \sigma_z(z_{it}) > 0$.

A central assumption in our setting is that production is risky. At time t the entrepreneur chooses labor ℓ_{it}, and its marginal product is $z_{it}dt + \sigma_y dB_{yit}$. The first term is expected productivity and the second is the shock that makes hiring labor risky. Expected productivity z_{it} evolves stochastically, following a diffusion driven by B_{zit}. We assume z_{it} remains in a bounded interval $[0, \bar{z}]$ and converges to a unique non-degenerate stationary distribution.

[2] The case $\rho_e = \rho_w$ is easy to analyze and does not change the main results. See Appendix E.

The resource constraints are[3]

$$c_t = \int_0^1 c_{it} di + c_{wt} = \int_0^1 z_{it}\ell_{it} di, \tag{5}$$

$$\ell_t = \int_0^1 \ell_{it} di. \tag{6}$$

2.2 The Competitive Equilibrium

There are competitive markets for the consumption good, labor, a risk-free bond in zero net supply, and entrepreneurs' equity. The consumption good is the numeraire, with price normalized to 1. Markets are incomplete: entrepreneurs must retain a fraction $\phi \in (0, 1]$ of their idiosyncratic risk. Budget constraints are

$$dn_{wt} = (r_t n_{wt} - c_{wt} + w_t \ell_t)dt, \tag{7}$$

$$dn_{it} = (r_t n_{it} - c_{it} + (z_{it} - w_t)\ell_{it})dt + (\phi\sigma_y)\ell_{it} dB_{yit}. \tag{8}$$

where n_{it} and n_{wt} are agents' bond positions, r_t is the interest rate, and w_t the wage. Entrepreneurs' inability to share idiosyncratic risk is reflected in their budget constraint: they must retain an exposure ϕ to the idiosyncratic risk in their production.[4] Both entrepreneurs and workers face natural borrowing constraints. For entrepreneurs, this means $n_{it} \geq 0$. Workers cannot borrow more than the present value of labor income, $n_{wt} \geq -\int_t^\infty e^{-\int_t^s r_u du} w_s \ell_s ds$. Bond market clearing requires

$$n_{wt} + \int_0^1 n_{it} di = 0. \tag{9}$$

[3] In equation (5) we are assuming a law of large numbers applies. That is, we focus on allocations such that $\int_0^1 \left(\int_0^t \ell_{is}\sigma_y dB_{yis} \right) di = 0$.

[4] Entrepreneurs sell, in a competitive market, a claim to a fraction $(1 - \phi)$ of their profits $dY_{it} - w_t\ell_{it}dt$ in exchange for payment $a_{it}dt$. Since the idiosyncratic risk can be diversified, the market prices this claim so $\mathbb{E}\left[(1 - \phi)(dY_{it} - w_t\ell_{it}dt) - a_{it}dt\right] = 0$. Thus, the entrepreneur's budget constraint is $dn_{it} = (r_t n_{it} - c_{it})dt + \phi(dY_{it} - w_t\ell_{it}dt) + a_{it}dt = (r_t n_{it} - c_{it} + (z_{it} - w_t)\ell_{it})dt + (\phi\sigma_y)\ell_{it}dB_{yit}$. Absent quantification of σ_y, it is without loss of generality to consider the $\phi = 1$ case, since ϕ and σ_y always appear together as $\phi\sigma_y$.

The *representative worker's problem* is to choose $(c_{wt} > 0, \ell_t \geq 0)$ to maximize his utility subject to his budget constraint and natural borrowing constraint, taking prices (r, w) and initial bonds n_{w0} as given. An *entrepreneur's problem* is to choose $(c_{it} > 0, \ell_{it} \geq 0)$ to maximize his utility subject to his budget constraint and natural borrowing constraint, taking prices (r, w) and initial bonds n_{i0} as given. Markets clear if equations 5 and 6 hold (with equation 9 then holding by Walras' Law).

For a given initial distribution of bonds n_{w0} and $n_{i0} > 0$ for $i \in [0, 1]$ satisfying equation (9), a *Competitive Equilibrium* is an allocation (c_w, ℓ_w) and (c_i, ℓ_i) for $i \in [0, 1]$, and market prices (r, w) such that the representative worker and each entrepreneur solve their problem and markets clear.

2.3 Characterizing the Competitive Equilibrium

Optimality Conditions. The representative worker's optimal allocation is characterized by an Euler equation and the first order condition for labor supply:

$$\frac{dc_{wt}}{c_{wt}} = (r_t - \rho_w)dt, \tag{10}$$

$$c_{wt}^{-1} w_t = \ell_t^{1/\psi}. \tag{11}$$

An entrepreneur's optimal allocation can be characterized by an Euler equation and the first order condition for labor demand:[5]

$$\frac{dc_{it}}{c_{it}} = (r_t - \rho_e + \sigma_{cit}^2)dt + \sigma_{cit} dB_{yit}, \tag{12}$$

$$z_{it} - w_t \leq \sigma_{cit} \phi \sigma_y \quad \text{(with equality if } z_{it} \geq w_t\text{)}, \tag{13}$$

[5] See Appendices A and B for details.

where,

$$\sigma_{cit} = \frac{\ell_{it}}{n_{it}}(\phi\sigma_y) = \frac{\ell_{it}}{c_{it}}(\rho_e\phi\sigma_y). \tag{14}$$

The variable σ_{cit} in equation (12) is the volatility of consumption growth of entrepreneur i and is at the center of our analysis. Here we have used the fact that with log preferences $c_{it} = \rho_e n_{it}$. His consumption is therefore locally exposed only to the idiosyncratic risk in his output B_{yit} (and not the idiosyncratic risk in his productivity B_{zit}, although they may be correlated). The σ_{cit}^2 term in the Euler equation captures the precautionary saving motive.

Hiring labor is like investing in a risky asset. The $\sigma_{cit}\phi\sigma_y \geq 0$ term in equation (13) is a risk premium on labor: it captures the covariation between the marginal utility of the entrepreneur and the fraction ϕ of the marginal product of labor that cannot be shared. Equation (14) uses the budget constraint to express σ_{cit} in terms of labor demand ℓ_{it}. Hiring more labor exposes the entrepreneur to more risk, while bonds provide self insurance. The demand for labor trades off the expected marginal profit from hiring labor, $z_{it} - w_t$, against increased exposure to risk.

Entrepreneurs with $z_{it} > w_t$ hire labor ($\ell_{it} > 0$), while less-productive entrepreneurs do not produce. Homothetic preferences and linear budget constraints result in labor demand that, when positive, is linearly increasing in productivity and in bonds, as can be seen by combining equations (13) and (14). Given wages, entrepreneurs hire less labor compared to the complete markets case. Even though production of the homogeneous good is linear, the most productive entrepreneur does not hire all the labor because that would expose him to too much risk.

The risk premium on labor creates labor wedges at the entrepreneur level that look like markups. For entrepreneurs who produce,

$$1 + \pi_{it} := \frac{1}{w_{it}/z_{it}} = 1 + \frac{\sigma_{cit}\phi\sigma_y}{w_t}. \tag{15}$$

That is, we define the risk markup, $1 + \pi_{it}$, to be the ratio of the price of the good to the unit labor cost, which is greater than one. The risk-adjusted marginal product of labor, $z_{it} - \sigma_{cit}\phi\sigma_y$, is equalized to the labor cost, w_t, for all entrepreneurs who have positive labor demand. Thus, there is no true markup in the sense that the risk-adjusted marginal cost is equal to the price of the good for all active entrepreneurs. At the aggregate level, however, risk markups show up as a labor wedge generating misallocation that reduces TFP.

$$Z_t := \frac{c_t}{\ell_t} = \int_0^1 z_{it} \frac{\ell_{it}}{\ell_t} di, \tag{16}$$

$$1 + \pi_t := \frac{Z_t}{w_t} = \int_0^1 \frac{z_{it}}{w_t} \frac{\ell_{it}}{\ell_t} di. \tag{17}$$

Aggregate productivity, Z_t, is the labor-weighted average idiosyncratic productivity and the aggregate risk markup, $1 + \pi_t$, is the labor-weighted average idiosyncratic risk markup.

Aggregation and a Separation Property. The environment with log preferences and linear production allows for easy aggregation and a powerful separation result. We can solve for the cross-sectional allocations separately from aggregate employment and output. Because labor demand is linear, when computing equilibrium allocations we do not need to keep track of the joint distribution of bonds n_{it} and productivity z_{it}. It is enough to keep track of the aggregate bond holdings (or equivalently, consumption) of entrepreneurs for each level of productivity. At time t, let η_{it} be the consumption share of entrepreneur i, η_{et} be the consumption share of all entrepreneurs, η_{wt} be the consumption share of the representative worker, and $\eta_t(z)$ be the consumption share of all entrepreneurs with productivity z:

$$\eta_{it} := c_{it}/c_t, \quad \eta_{et} := \int_0^1 \eta_{it} di, \quad \eta_{wt} := c_{wt}/c_t, \quad \eta_t(z) := \int_{\{z_{it}=z\}} \eta_{it} \, di. \tag{18}$$

Then, $\eta_{et} = \int_0^{\bar{z}} \eta_t(z) dz$ and by the resource constraint $\eta_{wt} = 1 - \eta_{et}$.

We can rewrite labor demand as

$$\ell_{it} = \sigma_{cit}\eta_{it}\frac{c_t}{\rho_e\phi\sigma_y}. \tag{19}$$

Combine optimality conditions (13) and (19) with the resource constraints to obtain expressions for the wage and aggregate productivity as a function of the entrepreneur consumption-share distribution. That is, we implicitly define the functions $w(\eta_t)$ and $Z(\eta_t)$ by

$$\frac{1}{\rho_e\phi\sigma_y} \times \int_w^{\bar{z}} z\eta_t(z)\frac{z-w}{\phi\sigma_y}dz = 1, \tag{20}$$

$$\left(\frac{1}{\rho_e\phi\sigma_y} \times \int_w^{\bar{z}} \eta_t(z)\frac{z-w}{\phi\sigma_y}dz\right)^{-1} = \frac{c_t}{\ell_t} = Z. \tag{21}$$

We also obtain an expression for the exposure to risk for each entrepreneur of productivity type z at each point in time as a function of the entrepreneur's productivity level and the consumption-share distribution. That is, we define the risk exposure function as

$$\sigma_c(z,\eta_t) := \frac{(z-w(\eta_t))^+}{\phi\sigma_y}. \tag{22}$$

We can use the goods resource constraint and the Euler equations for the worker and entrepreneurs to describe the evolution of η_t over time by a Kolmogorov forward equation (KFE) that only depends on η_t itself,

$$\begin{aligned}\partial\eta(z,t) = {} & \eta(z,t)\left[-\mu_z'(z) + \mu_\eta(z,\eta_t) + \sigma_z'^2(z) + \sigma_z(z)\sigma_z''(z) - \left(\sigma_z'(z)\sigma_c(z,\eta_t) + \sigma_z(z)\sigma_c'(z,\eta_t)\right)\lambda_{yz}\right] \\ & + \eta_z'(z,t)\left[-\mu_z(z) + 2\sigma_z(z)\sigma_z'(z) - \sigma_z(z)\sigma_c(z,\eta_t)\lambda_{yz}\right] \\ & + \eta_{zz}''(z,t)\left[\frac{1}{2}\sigma_z^2(z)\right]\end{aligned} \tag{23}$$

where $\mu_z(z), \sigma_z(z), \lambda_{yz}$ are primitives and $\mu_\eta(z,\eta_t) = \bar{\rho}_t - \rho_e + \sigma_c^2(z,\eta_t) - \int \eta_t(z)\sigma_c^2(z,\eta_t)dz$. See Appendix F for details of the KFE derivation.

Constructing the Equilibrium. This separation result allows us to compute an equilibrium in two steps. In step one, start with a given $\eta_{i0} > 0$ such that $\eta_{e0} < 1$ and compute $\eta_0(z)$. Then, solve the KFE forward to obtain the path of $\eta_t(z)$. Then use equations (20)–(22) to solve for the path of w_t, Z_t, and $\sigma_c(z, \eta_t)$. At this point we must verify that there exists a positive w_t that solves equation (20) using η_t at all t.[6]

Due to the separation result, step one does not involve aggregate output c_t, aggregate employment ℓ_t, or interest rates r_t. In step two, we determine these by using the representative worker's labor supply condition to solve for aggregate employment and the definition of aggregate TFP to compute aggregate output

$$(\overbrace{\eta_{wt} Z_t \ell_t}^{c_{wt}})^{-1} w_t = \ell_t^{1/\psi}, \tag{24}$$

$$c_t = Z_t \ell_t. \tag{25}$$

Once we have the path of ℓ and c we can construct the rest of the equilibrium. Worker consumption is $c_{wt} = \eta_{wt} c_t$, and their Euler equation determines the interest rate

$$r_t = \rho_w + \frac{\dot{c}_{wt}}{c_{wt}}. \tag{26}$$

Set $c_{i0} = \eta_{i0} c_0$ and $\sigma_{cit} = \sigma_c(z_{it}, \eta_t)$. We can use entrepreneurs' Euler equation (12) to determine their consumption c_{it} and then their labor demand FOC in equation (14) determines ℓ_{it}. We set the bonds of each entrepreneur to $n_{it} = c_{it}/\rho_e$ and the bonds of the worker to $n_{wt} = -\int_0^1 n_{it} di$.

Proposition 1. *Let $\mathcal{C} = (c_w, \ell, c_i, \ell_i, r, w)$ be as constructed by the above procedure with associated initial bonds n_{w0} and n_{i0}. Assume that for all t, $w_t, c_{it}, c_{wt} > 0$ and $\lim_{T\to\infty} e^{-\int_0^T r_u du} n_{wT} = 0$. Then $\mathcal{C}$ is a competitive equilibrium for initial bonds n_{w0} and n_{i0}.*

Proof. See Appendix C.

[6] If not, no competitive equilibrium exists that is consistent with initial condition n_{i0}, n_{w0}.

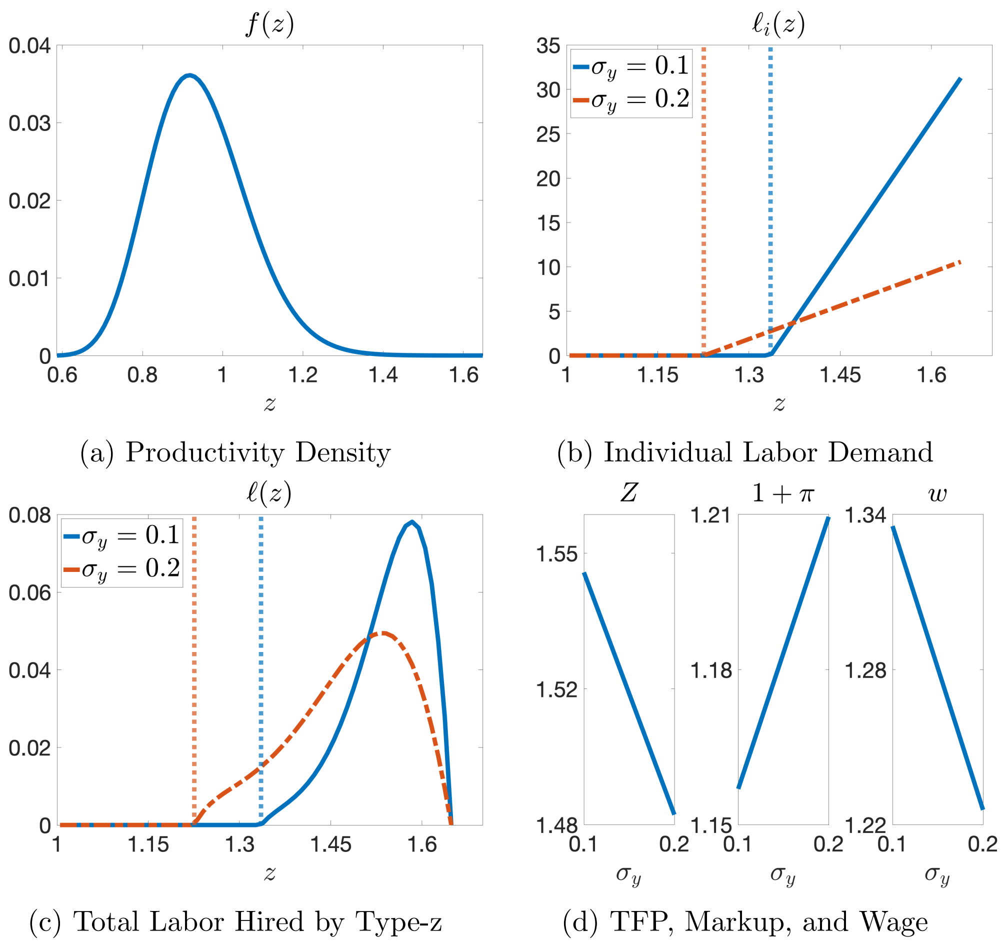

Figure 1: Comparative Static for σ_y

2.4 A Numerical Example: Equilibrium

To illustrate key characteristics of the competitive equilibrium, we explore comparative statics around some roughly chosen baseline parameter values. See Appendix G for details. Figure 1 explores how labor demand, TFP, and the aggregate markup change when hiring labor becomes more risky (increase in σ_y or, equivalently, ϕ). Blue solid lines are from the baseline calibration and orange dotted lines are from a counterfactual equilibrium with a larger σ_y. The vertical lines mark equilibrium wages.

Figure 1a plots the stationary productivity density. Figure 1b shows that an entrepreneur's labor demand is positive when productivity is higher than the wage and increases linearly in productivity (holding fixed wealth). When production becomes riskier, entrepreneurs do not scale their firm size with their productivity as aggressively. Figure 1c plots the distribution of total labor hired by all entrepreneurs of type z. Three forces determine the distribution of labor across productivity types. First, the density of z is decreasing for $z > w$. Second, the entrepreneur's optimal labor hiring policy is linearly increasing in productivity. Third, labor demand is a function of the bond distribution.

In this numerical example, there is not much labor hired by low productivity types close to the wage threshold even though there are many entrepreneurs of this type because they each hire little labor. Total labor hired by z-types initially increases in z even though the density of entrepreneurs is decreasing, because each entrepreneur is hiring more workers. Finally, as z further increases, the density of entrepreneurs falls fast enough that the total labor hired by z-type entrepreneurs falls, even as each entrepreneur continues to hire more workers.

Figure 1d shows the effect of increased risk on key macro aggregates. In an economy where production is riskier, each entrepreneur would hire less labor holding the wage constant. In equilibrium, the wage falls, bringing previously inactive entrepreneurs into production. The end result is that the largest producers reduce labor demand and the smallest producers increase labor demand. The aggregate markup rises, reflecting the increase in risk premia. In this model risk generates misallocation, lowering TFP. Connecting to the misallocation literature, more risk (higher σ_y) or less risk sharing (higher ϕ) could be a potential explanation for differences in cross-country TFP levels and the size distribution of firms. We now turn to analyzing the efficiency properties of the competitive equilibrium.

3 Optimal Policy

We consider a social planner who can use two policy instruments: (1) a uniform labor tax on the representative worker and (2) lump-sum transfers across agents at time $t = 0$. Motivated by asymmetric information concerns, we study uniform taxes that do not depend on idiosyncratic characteristics.[7] A tax on labor is the most policy-relevant instrument, and is useful for comparison to the literature, but we also consider an extension with taxes on savings. The $t = 0$ transfers can accomplish any redistribution, so taxes are used only for efficiency purposes. We also consider restrictions on the planner's ability to redistribute.

3.1 The Planner's Problem

The planner has access to a uniform labor tax τ_t that is rebated lump-sum to the worker. A tax plan is *feasible* if $w_t\tau_t$ is a bounded process. A labor tax rebated to workers distorts the labor margin without affecting any budget constraint in equilibrium. The representative worker's budget constraint is

$$dn_{wt} = (n_{wt}r_t - c_{wt} + (1 - \tau_t)w_t\ell_t + T_t)dt, \tag{27}$$

with natural borrowing limit $n_{wt} \geq -\int_t^\infty e^{-\int_t^s r_u du}((1 - \tau_s)w_s\ell_s + T_s)ds$. In equilibrium $T_t = \tau_t w_t \ell_t$. The equilibrium condition for labor supply becomes

$$\ell_t^{1/\psi} = c_{wt}^{-1}w_t(1 - \tau_t) = (\eta_{wt}Z_t\ell_t)^{-1}w_t(1 - \tau_t). \tag{28}$$

Agents' problems are unchanged, except that the representative worker's budget constraint is now given by equation (27) with the associated natural borrowing constraint. A *competitive equilibrium with uniform labor tax and lump-sum transfers* is an allocation (c_w, ℓ, c_i, ℓ_i), prices (r, w), taxes and transfers (τ, T) and initial bonds (n_{i0}, n_{w0}) such that (a) the representative worker and each entrepreneur solve their problem given their initial bonds, (b) markets clear, and (c) $T_t = \tau_t w_t \ell_t$ and $\int n_{i0}di + n_{w0} = 0$. An allocation is *implementable* if there exist prices, feasible taxes and transfers, and initial bonds such that they constitute a *competitive equilibrium with uniform labor tax and lump-sum transfers.* An implementable allocation is

[7] For example, firms can merge on paper to harvest size-dependent subsidies without changing any economic activity, or can misreport their activity (Chen, Liu, Serrato, and Xu, 2021).

optimal if it maximizes

$$\int_0^\infty e^{-\rho_w t}\tilde{\Gamma}_w \left(\log c_{wt} - \frac{\ell_t^{1+1/\psi}}{1+1/\psi}\right) dt + \int_0^1 \mathbb{E}\left[\int_0^\infty e^{-\rho_e t}\tilde{\Gamma}_e \log c_{it} dt\right] di, \tag{29}$$

where $\tilde{\Gamma}_w > 0$ and $\tilde{\Gamma}_e > 0$ are the Pareto weights on the representative worker and entrepreneurs. For simplicity of exposition, we assume that the planner does not care about entrepreneurs individually (i.e., $\tilde{\Gamma}_i = \tilde{\Gamma}_e \; \forall \; i$). It will be useful to work with the discounted Pareto weights,

$$\Gamma_{et} = e^{-\rho_e t}\tilde{\Gamma}_e, \quad \Gamma_{wt} = e^{-\rho_w t}\tilde{\Gamma}_w. \tag{30}$$

Transforming the Planner's Problem. In principle, the planner's problem is complicated by needing to choose an optimal transition path for an entire distribution. However, we can exploit a separation property to simplify the problem. This is done in two steps.

First, the objective function can be written to separate aggregate labor ℓ_t and the distribution of consumption shares η_{it}:[8]

$$\begin{aligned} &\overbrace{\int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log \ell_t - \Gamma_{wt}\frac{\ell_t^{1+1/\psi}}{1+1/\psi} dt}^{\text{labor}} + \overbrace{\int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log Z_t dt}^{\text{TFP}} \\ &+ \underbrace{\int_0^\infty \Gamma_{wt} \log \eta_{wt} + \Gamma_{et} \log \eta_{et} dt}_{\text{workers vs. entrepreneurs}} - \underbrace{\int_0^\infty \Gamma_{et}\left(\log \eta_{et} - \int_0^1 \log \eta_{it} di\right) dt}_{\text{entrepreneur inequality}}. \end{aligned} \tag{31}$$

Aggregate labor only appears in the first term of equation (31). The second term involves TFP. The third term captures the distribution of consumption between workers and entrepreneurs at the aggregate level. The fourth term reflects that the planner dislikes inequality among entrepreneurs (it is weakly positive due to Jensen's inequality).

Second, the choice of aggregate labor can also be separated from TFP and the distribution

[8]Write $c_{wt} = \eta_{wt} Z_t \ell_t$ and $c_{it} = \eta_{it} Z_t \ell_t$, apply logs, and use a law of large numbers to eliminate the expectation. Then add and subtract $\int_0^\infty \Gamma_{et} \log \eta_{et} dt$ and regroup terms.

of consumption shares in the constraints. Any implementable allocation must satisfy the resource constraints (5) and (6), the Euler equations (10) and (11), and the first order conditions for exposure to risk (14) and labor demand (19). Given these constraints, once the planner chooses the initial distribution of η_{i0}, η_t will then follow its KFE and pin down $Z(\eta_t)$, $w(\eta_t)$ and $\sigma_c(z, \eta_t)$ according to equations (20)–(23). Thus, the path of Z_t and η_{it} (and therefore η_{wt} and η_{et}) are independent of aggregate labor.

We can therefore recast the planner's problem as choosing an initial distribution η_{i0} and a path for aggregate labor ℓ to maximize the objective in equation (31) subject to the KFE constraint given by equations (20)–(23). See Appendix D for details on how to construct an optimal allocation from the solution to this problem. A key property of this formulation is that the choices of ℓ and the distribution of η_{i0} can be made independently, which delivers analytical tractability.

3.2 The First-Best Allocation

In the first-best allocation the planner optimizes each term in the objective function. (1) Aggregate labor is set to maximize the first term. (2) Labor is employed only by the highest productivity types. (3) Consumption is split between entrepreneurs and workers according to Pareto weights. (4) There is perfect risk sharing and all entrepreneurs have the same consumption. The first-best allocation is

$$\ell_t^{FB} = \left(\frac{1}{\frac{\Gamma_{wt}}{\Gamma_{wt}+\Gamma_{et}}} \right)^{\frac{\psi}{\psi+1}}, \tag{32}$$

$$Z_t^{FB} = \bar{z}, \tag{33}$$

$$\eta_{et}^{FB} = \frac{\Gamma_{et}}{\Gamma_{wt} + \Gamma_{et}}, \ \eta_{wt}^{FB} = \frac{\Gamma_{wt}}{\Gamma_{wt} + \Gamma_{et}}, \tag{34}$$

$$\eta_{it}^{FB} = \eta_{et}^{FB}. \tag{35}$$

The planner makes workers supply more labor when entrepreneurs' Pareto weight is larger.

Implementation of the First-Best Allocation. With complete markets the competitive equilibrium achieves the first best and the planner implements it with only redistribution at $t = 0$ and no taxes, $\tau_t = 0$.[9] Agents perfectly share risk, $\sigma_{cit} = 0$, so there are no risk markups, $w_t = \bar{z}$, and only the most productive entrepreneurs produce, $Z_t = \bar{z}$. Workers and entrepreneurs follow their Euler equations without precautionary saving or risk, so $\eta_{it} = \eta_{et}$, and η_{et} follows

$$d\eta_{et} = \eta_{et}(1 - \eta_{et})(\rho_e - \rho_w)dt. \tag{36}$$

This yields $\eta_{et} = \frac{e^{-\rho_e t}\eta_{e0}}{e^{-\rho_w t}\eta_{w0} + e^{-\rho_e t}\eta_{e0}}$. If the planner chooses the initial distribution $\eta_{e0} = \frac{\Gamma_{e0}}{\Gamma_{e0}+\Gamma_{w0}}$, we obtain $\eta_{et} = \frac{\Gamma_{et}}{\Gamma_{et}+\Gamma_{wt}}$. Aggregate labor would then be given by equation (11), which given w_t, Z_t, and η_{wt}, yields the first-best path for aggregate labor.

3.3 The Constrained Optimal Allocation

Since ℓ_t only appears in the labor term of the planner's objective, the optimal aggregate labor choice in the constrained planner problem coincides with the first-best, $\ell_t = \ell_t^{FB}$. Incomplete markets do not prevent the planner from obtaining the first-best aggregate labor.

In contrast, due to market incompleteness and the limited instruments that the planner has, the choice of the initial wealth distribution η_0 is distorted relative to the first best. The choice of η_0 determines the path of η_t according to the KFE in equation (23). The planner takes into account that the initial distribution of wealth η_0 affects both the path of misallocation, Z_t, and the consumption distribution.

Let's understand how the initial distribution of wealth can affect misallocation and risk sharing. First, TFP is a labor-weighted average of idiosyncratic productivity, and labor hired is increasing in an entrepreneur's wealth. Thus, the planner can increase TFP by allocating a larger wealth share to higher-productivity entrepreneurs. To the extent that the

[9] This is an application of the first and second welfare theorems.

wealth distribution is somewhat persistent, an initial redistribution can raise TFP along the transition.

Second, entrepreneurs must be exposed to risk if they are to employ labor, and this exposure to risk is higher the lower is their wealth share. This is costly for two reasons: (1) entrepreneurs utility is penalized because they are risk averse, and (2) the path of consumption shares of workers and entrepreneurs departs from the first-best due to entrepreneurs' precautionary saving, i.e., $\eta_{et} \neq \eta_{et}^{FB}$.

A Special Case with Only One Productivity Type. To see the second effect of how risk distorts the distribution of consumption more clearly, consider the special case with only one productivity type, $z_{it} = \bar{z}$. In this case, misallocation is not an issue, $Z_t = \bar{z}$. Furthermore, Appendix D shows that the last two terms of the objective function can be rewritten as

$$\underbrace{\int_0^\infty \Gamma_{wt} \log \eta_{wt} + \Gamma_{et} \log \eta_{et} dt}_{\text{workers vs. entrepreneurs}} - \underbrace{\int_0^\infty \Gamma_{et} \frac{1}{\rho_e} \frac{1}{2} \sigma_{ct}^2 dt}_{\text{entrepreneur inequality}} . \tag{37}$$

With complete markets, the planner would choose $\eta_{e0} = \frac{\Gamma_{e0}}{\Gamma_{e0}+\Gamma_{w0}}$, and agents would maximize the two terms on their own without further intervention. In the constrained problem, this is not possible. Equilibrium risk exposure is given by

$$\sigma_c(\eta_{et}) = \frac{\rho_e \phi \sigma_y}{\bar{z} \eta_{et}}. \tag{38}$$

Since the planner uses the labor tax to keep aggregate employment fixed at the first best, a lower wealth share for entrepreneurs η_{et} requires them to take more risk, σ_{ct}. This is costly because entrepreneurs are risk averse (entrepreneur inequality term) and also because their precautionary saving distorts the distribution of consumption (workers-vs.-entrepreneurs term).

Entrepreneurs' consumption share η_{et} follows the law of motion

$$d\eta_{et} = \eta_{et}(1 - \eta_{et})(\rho_w - \rho_e + \sigma_{ct}^2)dt. \tag{39}$$

Comparing this expression with equation (36), we see that the precautionary saving term, σ_{ct}^2, backloads entrepreneurs' consumption share relative to the path that maximizes the workers-vs.-entrepreneurs term. The result is stark when we consider the long run. In the first best $\eta_{et} \to 0$, while in the constrained optimal allocation

$$\eta_{et} \to \eta_e^* = \min\left\{\frac{1}{\sqrt{\rho_e - \rho_w}} \frac{\rho_e \phi \sigma_y}{\bar{z}}\ ,\ 1\right\}. \tag{40}$$

Even in this case without misallocation, the planner deviates from $\eta_{e0} = \eta_{e0}^{FB}$, trading off the welfare loss from precautionary saving (first term) and exposure to risk (second term).

3.4 Optimal Taxes

The constrained efficient planner achieves the first-best labor allocation by using a labor tax to correct the inefficient competitive equilibrium. The equilibrium labor supply condition in equation (28) can be expressed as

$$c_{wt}^{-1} \times \underbrace{Z_t \times \frac{1}{1+\pi_t}}_{w_t} \times (1 - \tau_t) = \ell_t^{1/\psi}. \tag{41}$$

The planner's first order condition can be expressed as[10]

$$c_{wt}^{-1} \times Z_t \times \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{wt}+\Gamma_{et}}} = \ell_t^{1/\psi}. \tag{42}$$

Comparison of the two equations directly reveals the optimal tax:

$$\boxed{1 - \tau_t = (1 + \pi_t) \times \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{wt}+\Gamma_{et}}}}. \tag{43}$$

[10]Isolate $\ell_t^{\frac{1}{\psi}}$ in equation (32), multiply and divide by $\eta_{wt} Z_t$, and use $c_{wt} = \eta_{wt} Z_t \ell_t$.

Assume that in the long run as $t \to \infty$ the consumption distribution converges to a stationary distribution η^*. Then $\pi(\eta_t) \to \pi^*$ and $\eta_w(\eta_t) \to \eta_w^*$ also converge to long-run values. The optimal tax converges to

$$1 - \tau^* = (1 + \pi^*) \times \eta_w^*. \tag{44}$$

We can write the optimal tax in the long run in terms of consumption shares as

$$1 - \tau^* = \frac{c_w^*}{w^* \ell^*}. \tag{45}$$

Entrepreneur wealth must be positive, $n_e^* = \frac{Z^* \ell^* \eta_e^*}{\rho_e} > 0$, so by market clearing $n_w^* < 0$, and thus the budget constraint of the worker implies $c_w^* < w^* \ell^*$. In the long run, workers are paying interest on their debt to entrepreneurs, so workers consume less than their labor income and the optimal labor tax is positive: $\tau^* > 0$.

We can further characterize the optimal path of the tax when $\eta_0 = \eta^*$. This is the tax the planner would choose if the economy started at steady state and he could not redistribute at time zero. Equation (43) simplifies to

$$1 - \tau_t = \frac{1 - \tau^*}{\frac{\Gamma_{wt}}{\Gamma_{wt} + \Gamma_{et}}}. \tag{46}$$

When the Planner does not value entrepreneurs welfare ($\tilde{\Gamma}_e \to 0$), we have $\tau_t = \tau^*$ for all t. When $\tilde{\Gamma}_e > 0$, τ_t is increasing over time. This could be a subsidy changing to a tax or a small tax growing to be a larger tax. As long as the planner does not care too much about entrepreneurs, if $\tilde{\Gamma}_e \leq \left(\frac{\tau^*}{1-\tau^*}\right) \tilde{\Gamma}_w$, then it is optimal to impose a tax along the entire transition path, i.e., $\tau_t \geq 0$ for all time.

3.5 Understanding the Inefficiency

The optimal labor tax in equation (43) can be understood in terms of two elements: inefficient risk premia and inefficient precautionary saving. Both arise from the single feature of incomplete idiosyncratic risk sharing. As a result, optimal policy when misallocation is caused by uninsurable risk is a tax, which is the opposite of an optimal subsidy when misallocation is caused by market power (in the Baqaee and Farhi (2020) and Edmond, Midrigan, and Xu (2023) sense). The remainder of the paper clarifies this insight.

Inefficient Risk Premia. The first term on the right of equation (43) is the aggregate risk markup, $1 + \pi_t$. The optimal labor tax has a component that aims to exactly undo the aggregate risk markup with a subsidy, as would be the optimal policy if the wedge was created by market power. Here, however, the wedge represents a risk premium that captures a real utility cost of exposing entrepreneurs to more risk. So why is it inefficient? The answer is that there is an externality associated with risk sharing. To see this, we can write an entrepreneur's exposure to risk in the following way:

$$\sigma_{cit} = (\rho_e \phi \sigma_y) \frac{\ell_{it}}{c_{it}} = (\rho_e \phi \sigma_y) \left(\frac{1}{\eta_{it} Z(\eta_t)} \right) \frac{\ell_{it}}{\ell_t}. \tag{47}$$

If an entrepreneur increases his labor demand ℓ_{it} individually, he is taking more risk. This is why in the competitive equilibrium entrepreneurs demand a risk premium as compensation. But the planner can implement a deviation that raises all entrepreneurs' ℓ_{it} at the same time and in the same proportion, so that the numerator ℓ_{it} and the denominator ℓ_t both increase proportionally and each entrepreneur's exposure to risk is unchanged. In terms of equilibrium objects, $\sigma_{cit} = (\phi \sigma_y) \frac{\ell_{it}}{n_{it}}$. An individual entrepreneur takes n_{it} as given when choosing ℓ_{it}. If the planner raises aggregate output c_t at time t by raising ℓ_t, then $n_{it} = c_{it}/\rho_e$ raises proportionally; this must occur because $\eta_{it} = c_{it}/c_t$ is predetermined. Since ℓ_{it} is linear in n_{it}, it also raises in proportion. Thus, σ_{cit} is unchanged. In the background, the interest rate increases, according to the Euler equations. Entrepreneurs enter period t with higher n_{it} because the past rate of return on their savings (the interest rate) increases with this deviation. Thus, they have more self-insurance to smooth consumption in the face of the increased risk associated with hiring more labor. The risk premium correctly captures the private cost of exposing entrepreneurs to risk at the margin, but not the correct social cost.

Inefficient Precautionary Saving. The second term in equation (43) is

$$\frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{wt}+\Gamma_{et}}} \tag{48}$$

and captures the cumulative effects of inefficient precautionary saving. The numerator is the share of consumption that goes to the representative worker and the denominator is their relative Pareto weight. In the absence of incomplete risk sharing (or in the first best), this ratio would always be one. With incomplete risk sharing, this is not the case due to entrepreneurs' precautionary saving, as shown above.

But even if entrepreneurs' share of consumption is too high, why does this show up as an inefficiency in the aggregate level of labor? The reason is that, from the perspective of the planner, if he increases labor and output at time t, the share of the extra output that goes to entrepreneurs above their Pareto weight is wasted. So he acts as if the marginal product of labor was not Z_t but rather Z_t times this ratio—as can be seen clearly in the first order condition in equation (42). In terms of equilibrium objects, workers take their wealth $n_{wt} < 0$ as given when choosing their labor supply. The planner realizes that reducing aggregate employment ℓ_t reduces workers' debt to entrepreneurs (shrinks n_{wt} towards zero), so more of their income goes to their consumption instead of debt payments to finance entrepreneur consumption. This improves the Planner's objective to the extent that the consumption distribution does not match Pareto weights due to entrepreneurs' past precautionary saving.

3.5.1 A Tax on Entrepreneurs' Savings

To understand the inefficiency of precautionary saving, consider what the planner would do if he could also tax entrepreneurs' savings.[11] To highlight the role of risk sharing, we return to the simplified environment with only one productivity type, $z_{it} = \bar{z}$. This eliminates the planner's concern for misallocation. As in equation (37), the objective function, omitting the independent labor term, reduces to

$$\underbrace{\int_0^\infty \Gamma_{wt} \log \eta_{wt} + \Gamma_{et} \log \eta_{et} dt}_{\text{workers vs. entrepreneurs}} - \underbrace{\int_0^\infty \Gamma_{et} \frac{1}{\rho_e} \frac{1}{2} \sigma_{ct}^2 dt}_{\text{entrepreneur inequality}} .$$

Risk exposure is as in equation (38),

$$\sigma_c(\eta_{et}) = \frac{\rho_e \phi \sigma_y}{\bar{z} \eta_{et}}.$$

The law of motion of η_{et} is similar to equation (39), but now contains the savings tax τ_{st},

$$d\eta_{et} = \eta_{et}(1 - \eta_{et})(\rho_w - \rho_e + \sigma_{ct}^2 - \tau_{st})dt. \tag{49}$$

With this new instrument, entrepreneurs' consumption share becomes a control instead of a state variable. The workers-vs.-entrepreneurs term is maximized when consumption shares equal Pareto weights, but the planner chooses to give entrepreneurs a larger consumption share to reduce their exposure to risk, and thus entrepreneur inequality. The FOC for η_{et} is

$$\Gamma_{et} \frac{1}{\eta_{et}} - \Gamma_{wt} \frac{1}{1 - \eta_{et}} + \Gamma_{et} \frac{1}{\rho_e} \sigma_c^2(\eta_{et}) \frac{1}{\eta_{et}} = 0.$$

[11] We think that this policy instrument that treats entrepreneurs and workers differently is less realistic than the labor tax. The more realistic uniform savings tax is a redundant policy instrument as long as the planner has access to $t = 0$ lump-sum transfers. All agents have the same interest rate in their Euler equations, and a subsidy/tax that affects the return on savings equally for all agents simply serves to redistribute without affecting efficiency.

Re-arranging,[12]

$$\eta_{et} = \frac{\Gamma_{et}}{\Gamma_{wt} + \Gamma_{et}} \times \left(1 + (1 - \eta_{et}) \frac{1}{\rho_e} \sigma_c^2(\eta_{et})\right). \tag{50}$$

From this we deduce that entrepreneurs' consumption share η_{et} falls over time, but is always above their Pareto weight. In the long run, it converges to zero, $\eta_{et} \to 0$. In contrast, without the tax on savings, $\eta_{et} \to \eta_e^*$ from equation (40).

How is this allocation implemented? The optimal labor tax is still given by equation (43). To implement the optimal path for η_{et}, the saving tax must equate equation (49) with the law of motion of equation (50). After some algebra, we obtain

$$\tau_{st} = \sigma_{ct}^2 + (\rho_w - \rho_e) \left(1 - \frac{\frac{\frac{\Gamma_{wt}}{\Gamma_{wt} + \Gamma_{et}}}{\eta_{wt}}}{1 + \frac{\frac{\Gamma_{et}}{\Gamma_{wt} + \Gamma_{et}}}{\eta_{et}} \frac{1}{\rho_e} \sigma_c^2(\eta_{et}) \left(2 - \eta_{et}\right)}\right). \tag{51}$$

The tax on savings first eliminates precautionary saving, σ_{ct}^2, but then nonetheless distorts consumption paths relative to the first best to account for the effect of η_{et} on risk sharing.

In the long run, entrepreneurs' consumption share η_{et} converges to zero, so their exposure to risk $\sigma_c(\eta_{et})$ goes to infinity, as do risk markups, $1 + \pi_t$, and precautionary saving, $\sigma_c^2(\eta_{et})$. In the limit, optimal taxes satisfy[13]

$$\lim_{t \to \infty} \frac{1 - \tau_t}{1 + \pi_t} = \lim_{t \to \infty} \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{wt} + \Gamma_{et}}} = 1, \tag{52}$$

$$\lim_{t \to \infty} \frac{\tau_{st}}{\sigma_{ct}^2} = 1. \tag{53}$$

The tax on labor converges to a subsidy (negative tax) to eliminate the aggregate risk markup, while the tax on savings converges to a tax to eliminate the precautionary saving motive. This is in contrast to the case without the tax on savings, where the tax on la-

[12] This is a cubic equation for η_{et}, so it has a closed-form expression, but it's not very nice, and this expression is more illuminating.

[13] For τ_{st}, use equation (50) to replace $\frac{\frac{\Gamma_{et}}{\Gamma_{wt}+\Gamma_{et}}}{\eta_{et}}$ in equation (51), divide by σ_{ct}^2 throughout, and take limits.

bor converges to a tax. The difference is precisely the behavior of η_{et}, which encodes the cumulative effect of precautionary saving on the distribution of consumption.

Another special case worth highlighting is the limit as $\tilde{\Gamma}_e \to 0$. From equation (50), we obtain $\eta_{et} \to 0$ for all t. The taxes that implement the optimal allocation satisfy, for all t,

$$\frac{1-\tau_t}{1+\pi_t} = \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{wt}+\Gamma_{et}}} \to 1, \tag{54}$$

$$\frac{\tau_{st}}{\sigma_{ct}^2} \to 1. \tag{55}$$

The planner uses the labor tax to exactly undo the markup and the tax on savings to exactly undo the precautionary saving, not just in the long run, but along the whole path.

In summary, without a tax on saving, the optimal labor tax deviates from $\tau_t = -\pi_t$ due to the cumulative effects of entrepreneurs' inefficient precautionary saving. If the planner has access to a tax on entrepreneurs' savings, he uses it to eliminate the precautionary motive, but still distorts the path of consumption shares relative to the first best to account for the effect of η_{et} on entrepreneurs' exposure to risk. In the long run (or when he doesn't care about entrepreneurs, $\tilde{\Gamma}_e \to 0$), the planner uses a negative tax (i.e., a subsidy) on labor to exactly offset the risk markup and a tax on saving to exactly offset precautionary saving.

3.6 Numerical Examples: Planner Problem

To illustrate key characteristics of the planner's allocation and implementation, we study the effects of different policies in both the long run and on the transition path, using a roughly chosen set of baseline parameter values. See Appendix G for details.

3.6.1 The Long Run

We first analyze how the planner's allocation differs from the zero-tax competitive equilibrium in the long run. In our numerical example, the long-run optimal labor tax is $\tau^* = 1-(1+\pi^*)\eta_w^* = 9\%$. The first row of Table 1 presents the percent change in long-run consumption,

Labor Tax: Long Run				
	τ	c	ℓ	Welfare
Optimal Tax: $\tau = \tau^*$	9%	-4%	-4%	0.21%
Markup Subsidy: $\tau = -\pi^*$	-16%	7%	7%	-1.28%

Table 1: The labor tax rate and the percent change in consumption, labor, and consumption-equivalent welfare induced by the tax compared to the zero-tax equilibrium. The first row is for the optimal long-run tax. The second is for a labor subsidy equal to the average markup.

labor, and consumption-equivalent welfare induced by the optimal labor tax. The tax reduces consumption and labor by 4% and increases welfare. According to the planner's objective, the welfare gain from the optimal tax is equivalent to increasing consumption in the zero-tax competitive equilibrium by 0.21% for each person and point in time.

To contrast our findings with those typical of the market-power markup literature, the second row of Table 1 compares the zero-tax competitive equilibrium to the allocation achieved by a labor subsidy equal to the average markup (i.e., $\tau = -\pi^*$). In the long run, the labor subsidy increases consumption and labor by 7% and reduces welfare by 1.28%.

3.6.2 Transition Dynamics

Results along the transition path depend on the values of the Pareto weights Γ_{wt} and Γ_{et}. For illustrative purposes, we set $\tilde{\Gamma}_e = 0.04$ and $\tilde{\Gamma}_w = 0.96$, motivated by evidence that approximately 4% of US households are entrepreneurs (Salgado, 2020). An interval $\Delta t = 1$ is one year. The transition differs from the long run because the planner can choose the initial distribution of consumption shares and because the discounted Pareto weights change over time inducing a time-varying path for labor. We first isolate the effect of choosing the path of labor and then study the effect of initial redistribution.

Fixed Initial Distribution: Choose ℓ_t given η_0. Figure 2 plots the optimal labor tax, aggregate labor, and consumption that solve the planner problem starting from the steady state of the competitive equilibrium in $t = 0$. By the separation result, the distribution of consumption shares and aggregate productivity are identical to the competitive equilibrium, and

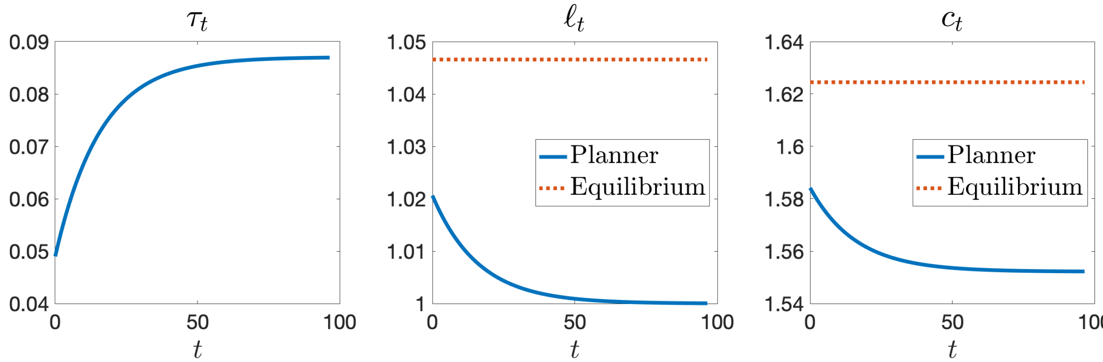

Figure 2: Optimal labor tax τ_t, aggregate labor ℓ_t, and consumption c_t that solve the planner problem, taking as given that $t = 0$ is the steady state of the zero-tax equilibrium.

Labor Tax: Transition Dynamics			
	$\tau_t = 1 - (1 + \pi_t)(\eta_{wt})$	$\tau_t = \tau^*$	$\tau_t = -\pi_t$
Welfare gain	0.15%	0.14%	-1.15%

Table 2: We report the labor tax and the associated change in consumption-equivalent welfare compared to zero taxes. The first column implements the optimal dynamic tax. The second column implements a constant tax at the optimal long-run tax rate. The third column implements a dynamic labor subsidy equal to the average markup.

changes in welfare must come from changes in the labor term of the transformed planner's objective:

$$\int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log \ell_t - \Gamma_{wt} \frac{\ell_t^{1+1/\psi}}{1 + 1/\psi} dt. \tag{56}$$

The key difference from the long run is that the planner values entrepreneurs' consumption on the transition. This implies that the optimal labor tax at $t = 0$ is less than its long run value, temporarily increasing aggregate labor and consumption. The first column of Table 2 reports a 0.15% welfare gain from the optimal dynamic labor tax τ_t. The second column shows that the welfare gain from implementing a constant labor tax equal to the long-run optimal rate is 0.14%, very close to that from implementing the optimal dynamic tax. Finally, the third column shows that implementing a dynamic labor subsidy equal to the average markup leads to a 1.15% consumption equivalent welfare loss. Thus, in our exercise, long-run results are a good guide to optimal policy on the transition and mistakenly implementing the optimal policy associated with market-power markups is very costly.

Initial Redistribution: Choose ℓ_t and η_0. We now study a planner who uses labor taxes and can redistribute at time zero. Figure 3 plots the transition dynamics of this economy in comparison to the economy in which the planner cannot initially redistribute. In the long run, entrepreneurs have a consumption share above 20% while their discounted Pareto weight converges to zero. Even though along the transition Γ_{et} is not zero, it is still well-below η_e^*, and so the planner wants to initially redistribute consumption to the workers. In this example, the planner sets $\eta_{e0} = 0.037$, close to the first-best value of $\tilde{\Gamma}_e = 0.04$. The planner always sets labor to its first best level. After the initial redistribution, workers' consumption share is above their Pareto weight and the income effect reduces their labor supply, so the planner needs to subsidize labor in the short run. As the entrepreneurs accumulate wealth over time and $\eta_t \to \eta^*$, the labor subsidy becomes a tax along the transition.

TFP and the aggregate markup converge quickly to their long-run values, even though the consumption share of entrepreneurs is much slower to converge. To better understand TFP and risk markup dynamics, in Figure 3 we plot the distribution of consumption shares among entrepreneurs $\eta_t(z)$ and the total labor hired by type-z entrepreneurs $\ell_t(z)$ at different times along the transition. The key insight is that TFP and risk premia are determined by the wealth shares of entrepreneurs who have high-enough productivity to produce, while the majority of entrepreneurs' wealth is held by lower-productivity individuals. In this example, the majority of entrepreneurs are rentiers, low productivity and not actively producing, living off the profits of days past and waiting to become profitable again in the future.

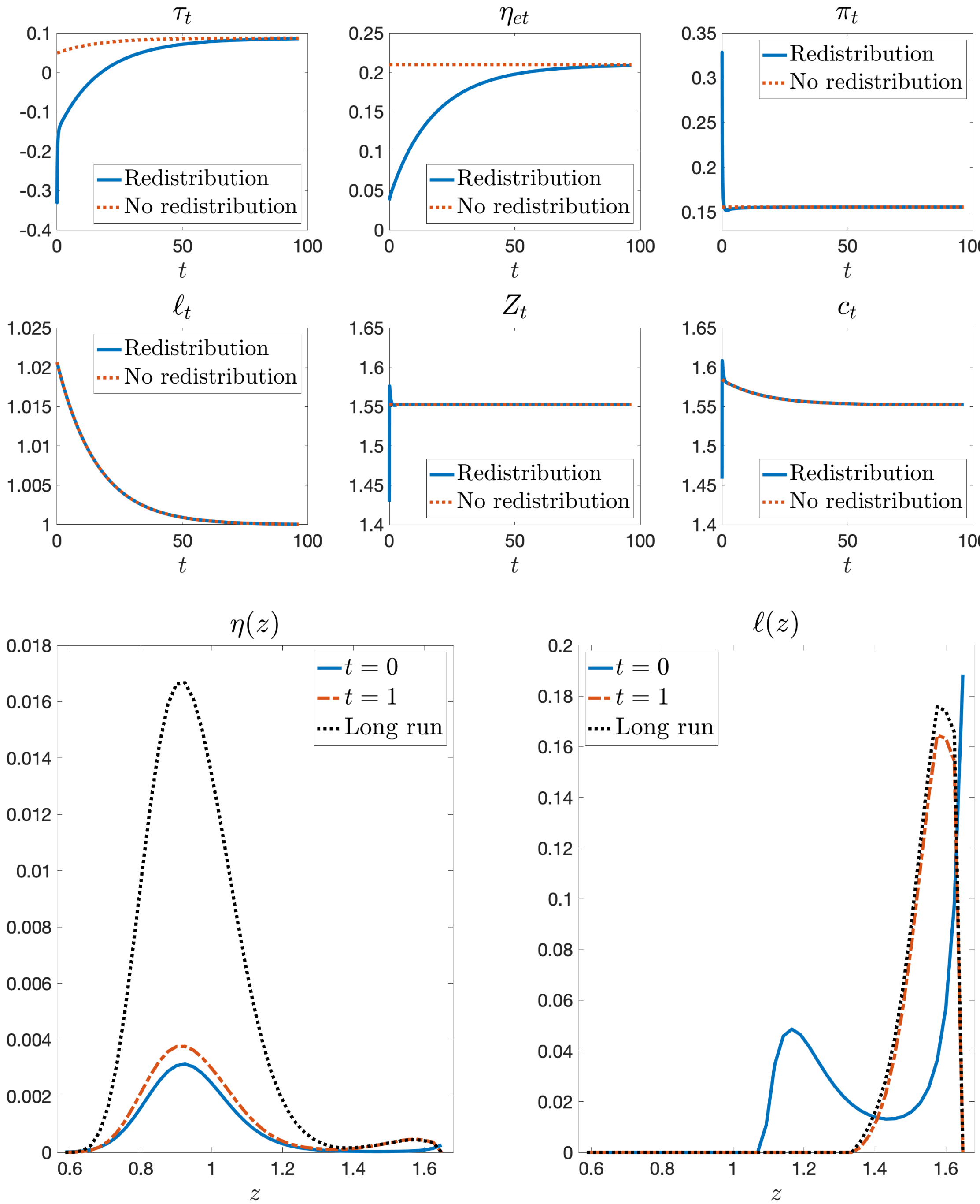

Figure 3: We plot the optimal labor tax τ_t, the consumption share of entrepreneurs η_{et}, the aggregate risk markup π_t, TFP Z_t, and aggregate consumption c_t in the planner's allocation. The planner chooses the labor tax and how to redistribute wealth between entrepreneurs and workers at $t = 0$, starting from the steady state of the competitive equilibrium. We also plot the distribution of the total consumption share of type-z entrepreneurs $\eta_t(z)$ and the total labor hired by type-z entrepreneurs $\ell_t(z)$ at different times along the transition.

Labor Tax and Redistribution: Transition Dynamics				
	Welfare Gain Decomposition			
Total	Labor	TFP	Workers vs. Entrepreneurs	Entrepreneur Inequality
5.69%	0.15%	0.03%	7.00%	-1.50%

Table 3: We report changes in welfare between the planner solution and the zero-tax competitive equilibrium. We consider the case in which the planner can redistribute wealth at $t = 0$ and implement the optimal dynamic labor tax. The first column reports the total consumption-equivalent welfare gain. The other four columns decompose changes in welfare according to components of the transformed planner's problem, reflecting changes in labor, aggregate TFP, inequality between workers and entrepreneurs, and inequality among entrepreneurs.

The planner initially lowers the consumption share of almost all entrepreneurs except for the most productive, who receive an increase. The right tail of the consumption share distribution converges quickly, limiting the effectiveness of targeted initial transfers.[14] This redistribution affects TFP and risk markups in two ways. The decline in the average consumption share of entrepreneurs increases risk markups and lowers TFP. Redistribution in favor of the highest-productivity entrepreneurs offsets this. Here, the first effect initially dominates, with a drop in TFP and increase in the aggregate risk markup. Then there is overshooting, as the wealth share of medium-productivity producers increases and drives out lower-productivity producers, before rapid convergence to the steady state.

Table 3 reports that the planner achieves a consumption-equivalent welfare gain of 5.5% and decomposes this gain into the four components of the planner's objective. In line with the separation result, the welfare gain from the labor component is identical to the case in which the planner cannot redistribute. However, the overall welfare gain is larger when the planner can redistribute, almost entirely due to improvements in the worker-vs.-entrepreneur consumption share component. There is also a decrease in welfare from the increased inequality among entrepreneurs caused by the increase in risk.

[14] An alternative experiment in which the planner who starts with the stationary distribution of wealth and can redistribute between entrepreneurs and workers by changing the consumption share of all entrepreneurs by the same proportion yields very similar results, other than an initially larger drop in TFP.

4 Conclusion

We study optimal policy when measured markups reflect compensation for risk. Hiring labor is like investing in a risky asset, so entrepreneurs trade off expected profits against uninsurable risk and demand risk premia. These risk premia create labor wedges at the entrepreneur level that look like measured markups. They lead to lower TFP (misallocation), an aggregate labor wedge, and inequality both among entrepreneurs and between entrepreneurs and workers.

The constrained planner implements the first-best level of employment but cannot avoid distortions to the first-best consumption distribution because of uninsurable risk and precautionary saving. The optimal labor tax has a keep rate equal to the product of (1) the aggregate risk markup and (2) workers' consumption share relative to their Pareto weight. Although risk markups correctly reflect the private cost of risk, they do not capture the correct social cost. Inefficient risk premia call for a labor subsidy to eliminate the aggregate markup, just as the market-power perspective indicates.

However, precautionary saving generates an excessive concentration of wealth in the hands of entrepreneurs, and impoverished workers supply too much labor. Inefficient precautionary saving calls for a labor tax to offset this excessive labor supply. In the long run, the precautionary inefficiency dominates and the optimal policy is a tax—the opposite of an optimal subsidy to counteract market power. Alternatively, if the planner can also tax entrepreneurs' and workers' savings at different rates, he uses a labor subsidy to undo the aggregate markup and a entrepreneur savings tax to eliminate their precautionary saving. Although we consider the entrepreneur-specific savings tax less realistic, in both cases the policy recommendation is sharply different than that from the market-power perspective.

According to our simple numerical exercise, there is a modest welfare gain from implementing the optimal dynamic labor tax; almost the same gain from implementing a constant labor tax at the long-run optimal level; large welfare losses from implementing a labor subsidy equal to the aggregate markup; and a large welfare gain from redistributing from entrepreneurs to workers.

References

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

Algan, Y., O. Allais, and W. J. Den Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32(3), 875–908.

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

Asker, J., A. Collard-Wexler, and J. De Loecker (2014): "Dynamic Inputs and (Mis)Allocation," *Journal of Political Economy*, 122(5), 1013–1063.

Baqaee, D. R., and E. Farhi (2020): "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 135(1), 105–163.

Bloom, N., F. Guvenen, and S. Salgado (2023): "Skewed Business Cycles," *working paper*.

Boar, C., D. Gorea, and V. Midrigan (2023): "Why Are Returns to Private Business Wealth So Dispersed?," *American Economic Review*, R&R.

Buera, F. J., and Y. Shin (2013): "Financial Frictions and the Persistence of History: A Quantitative Exploration," *Journal of Political Economy*, 121(2), 221–272.

Chen, Z., Z. Liu, J. C. S. Serrato, and D. Y. Xu (2021): "Notching R&D Investment with Corporate Income Tax Cuts in China," *American Economic Review*, 111(7), 2065–2100.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review*, 101(3), 471–75.

David, J., R. Ranciere, and D. Zeke (2023): "Risk-Taking, Capital Allocation and Monetary Policy," *working paper.*

David, J., L. Schmid, and D. Zeke (2022): "Risk-Adjusted Capital Allocation and Misallocation," *Journal of Financial Economics*, 145(3), 684–705.

David, J., and D. Zeke (2024): "International Diversification, Reallocation, and the Labor Share," *working paper.*

De Loecker, J., and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102(6), 2437–2471.

Di Tella, S., and R. Hall (2022): "Risk Premium Shocks Can Create Inefficient Recessions," *The Review of Economic Studies*, 89(3), 1335–1369.

Dou, W. W., Y. Ji, D. Tian, and P. Wang (2025): "Misallocation and Asset Prices," *working paper.*

Edmond, C., V. Midrigan, and D. Y. Xu (2023): "How Costly Are Markups?," *Journal of Political Economy*, 131(7), 1619–1675.

Eeckhout, J., and L. Veldkamp (2023): "Data and Markups: A Macro-Finance Perspective," *working paper.*

Hall, R. E. (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96(5), 921–47.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Itskhoki, O., and B. Moll (2019): "Optimal Development Policies with Financial Frictions," *Econometrica*, 87(1), 139–173.

Khan, A., and J. K. Thomas (2013): "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity," *Journal of Political Economy*, 121(6), 1055–1107.

Knight, F. H. (1921): *Risk, Uncertainty, and Profit.* Houghton Mifflin Company.

Meh, C. A., and V. Quadrini (2006): "Endogenous Market Incompleteness with Investment Risks," *Journal of Economic Dynamics and Control*, 30(11), 2143–2165.

Menzio, G. (2024): "Markups: A Search-Theoretic Perspective," *working paper.*

Moll, B. (2014): "Productivity Losses from Financial Frictions: Can Self-financing Undo Capital Misallocation?," *The American Economic Review*, 104(10), 3186–3221.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11(4), 707–720.

Salgado, S. (2020): "Technical Change and Entrepreneurship," *working paper.*

A Entrepreneur Optimality

The entrepreneurs' problem is to chose adapted processes $(c_i > 0, \ell_i \geq 0)$ to maximize utility

$$U(c_i) = \mathbb{E}\left[\int_0^\infty e^{-\rho_e t} \log c_{it} dt\right] \tag{A.1}$$

subject to the dynamic budget constraint

$$dn_{it} = (r_t n_{it} - c_{it} + \ell_{it}(z_{it} - w_t))dt + \ell_{it}\sigma_y \phi dB_{yit} \tag{A.2}$$

with initial bonds n_{i0} given and the natural borrowing limit $n_{it} \geq 0$.

Any plan (c_i, ℓ_i) satisfying the dynamic budget constraint satisfies the intertemporal budget constraint

$$\mathbb{E}\left[\int_0^\infty \xi_{it} c_{it} dt\right] \leq n_{i0}, \tag{A.3}$$

where ξ_{it} is the solution to

$$d\xi_{it}/\xi_{it} = -r_t dt - \frac{(z_{it} - w_t)^+}{\phi\sigma_y} dB_{yit}, \tag{A.4}$$

with $\xi_{i0} = 1$. To see this, use Itô's lemma to compute

$$\begin{aligned}
d(\xi_{it} n_{it}) &= \xi_{it}(r_t n_{it} - c_{it} + \ell_{it}(z_{it} - w_t) - r_t n_{it} - \frac{(z_{it} - w_t)^+}{\phi\sigma_y}\ell_{it}\phi\sigma_y)dt + \xi_{it}\left(\ell_{it}\phi\sigma_y - \frac{(z_{it} - w_t)^+}{\phi\sigma_y} n_{it}\right) dB_{yt}, \\
&= \xi_{it}(-c_{it} + \ell_{it}\left((z_{it} - w_t) - (z_{it} - w_t)^+\right))dt + \xi_{it}\left(\ell_{it}\phi\sigma_y - \frac{(z_{it} - w_t)^+}{\phi\sigma_y} n_{it}\right) dB_{yt}.
\end{aligned}$$

Integrate and take expectations to obtain

$$\mathbb{E}\left[\xi_{i\tau_j} n_{i\tau_j}\right] = \xi_{i0} n_{i0} + \mathbb{E}\left[\int_0^{\tau_j} \xi_{it}(-c_{it} + \ell_{it}\left((z_{it} - w_t) - (z_{it} - w_t)^+\right))dt\right],$$

where $\{\tau_j\}$ is an increasing sequence of stopping times with $\lim_{j\to\infty} \tau_j = \infty$ a.s., such that

$\int_0^{\tau_j} \xi_{it} \left(\ell_{it}\phi\sigma_y - \frac{(z_{it}-w_t)^+}{\phi\sigma_y} n_{it} \right) dB_{yt}$ is a martingale. Reorganizing we get

$$\mathbb{E}\left[\int_0^{\tau_j} \xi_{it}(c_{it} - \ell_{it}\left((z_{it} - w_t) - (z_{it} - w_t)^+\right))dt\right] + \mathbb{E}\left[\xi_{i\tau_j} n_{i\tau_j}\right] = \xi_{i0} n_{i0}.$$

Notice that $\xi_{it}(c_{it} - \ell_{it}((z_{it} - w_t) - (z_{it} - w_t)^+)) \geq \xi_{it} c_{it} \geq 0$. Take limits as $j \to \infty$ using the monotone convergence theorem and recall $\xi_{it} n_{it} \geq 0$ for any plan that satisfies the natural borrowing constraint. We obtain

$$\mathbb{E}\left[\int_0^{\infty} \xi_{it} c_{it} dt\right] \leq n_{i0}.$$

We will therefore consider a relaxed problem: choose consumption c_i to maximize utility $U(c_i)$ subject to the intertemporal budget constraint (A.3) and then later check that the proposed solution satisfies the dynamic budget constraint. We have a first-order condition

$$e^{-\rho_e t} c_{it}^{-1} = \lambda \xi_{it}, \tag{A.5}$$

where the value of $\lambda > 0$ ensures the constraint holds with equality:

$$\begin{aligned} \mathbb{E}\left[\int_0^{\infty} \xi_{it} c_{it} dt\right] &= \mathbb{E}\left[\int_0^{\infty} e^{-\rho_e t} \lambda^{-1} dt\right] = \frac{1}{\rho_e \lambda} = n_{i0}, \\ \implies \lambda =& \frac{1}{\rho_e} \frac{1}{n_{i0}}. \end{aligned}$$

These conditions are sufficient for optimality of the relaxed problem. To see this, consider any alternative plan $\tilde{c}_i$ satisfying the intertemporal budget constraint. Use the gradient inequality for concave functions to note that $\log \tilde{c}_{it} \leq \log c_{it} + c_{it}^{-1}(\tilde{c}_{it} - c_{it})$. Then,

$$\begin{aligned}
U(\tilde{c}_i) &= \mathbb{E}\left[\int_0^\infty e^{-\rho_e t} \log \tilde{c}_{it} \; dt\right] \\
&\leq \mathbb{E}\left[\int_0^\infty e^{-\rho_e t} \left(\log c_{it} + c_{it}^{-1}(\tilde{c}_{it} - c_{it}\right) dt\right] \\
&= U(c_i) + \mathbb{E}\left[\int_0^\infty \lambda \xi_{it}(\tilde{c}_{it} - c_{it}) dt\right] \leq U(c_i).
\end{aligned}$$

The final inequality holds because $\mathbb{E}\left[\int_0^\infty \xi_{it}\tilde{c}_{it}\right] \leq n_{i0}$ since $\tilde{c}$ satisfies the intertemporal budget constraint in equation (A.3) while $\mathbb{E}\left[\int_0^\infty \xi_{it} c_{it}\right] = n_{i0}$ by construction.

We now want to derive the stochastic processes that these sufficient conditions imply. From the first-order condition (A.5), we obtain that $c_{it} > 0$ is an Itô process,

$$dc_{it}/c_{it} = \mu_{cit} dt + \sigma_{cit} dB_{yit}.$$

Using Itô's lemma we obtain

$$d(e^{-\rho_e t} c_{it}^{-1}) = d(\lambda \xi_{it}) \tag{A.6}$$

$$e^{-\rho_e t} c_{it}^{-1} \left(-\rho_e - \mu_{cit} + \sigma_{cit}^2\right) dt - e^{-\rho_e t} c_{it}^{-1} \sigma_{cit} dB_{yit} = -\lambda \xi_{it} r_t dt - \lambda \xi_{it} \frac{(z_{it} - w_t)^+}{\phi \sigma_y} dB_{yit}. \tag{A.7}$$

Matching terms we get the Euler equation and the first-order condition for labor,

$$\mu_{cit} = r_t - \rho_e + \sigma_{cit}^2, \tag{A.8}$$

$$\sigma_{cit} = \frac{(z_{it} - w_t)^+}{\phi \sigma_y}. \tag{A.9}$$

And any plan $c_i > 0$ satisfying these conditions and $c_{i0} = \lambda^{-1} = \rho_e n_{i0}$ satisfies FOC (A.5) and the intertemporal budget constraint with equality and is therefore optimal in the relaxed problem.

It only remains to show that our candidate plan c_i can indeed be implemented as part of a plan (c_i, ℓ_i) that satisfies the dynamic budget constraint. We set

$$\ell_{it} = \frac{\sigma_{cit} c_{it}}{\phi \sigma_y \rho_e} \geq 0, \tag{A.10}$$

We now define the process for bonds $n_{it} = c_{it}/\rho_e > 0$ which satisfies the initial condition by construction. Using the optimality conditions we compute

$$\begin{aligned} dn_{it}/n_{it} &= \left(r_t - \rho_e + \sigma_{cit}^2\right) dt + \sigma_{cit} dB_{yit}, \\ &= \left(r_t - c_{it}/n_{it} + \frac{\ell_{it}}{n_{it}} \phi \sigma_y \frac{(z_{it} - w_t)^+}{\phi \sigma_y}\right) dt + \frac{\ell_{it}}{n_{it}} \sigma_y \phi dB_{yit}, \\ &= \left(r_t - c_{it}/n_{it} + \frac{\ell_{it}}{n_{it}} (z_{it} - w_t)\right) dt + \frac{\ell_{it}}{n_{it}} \phi \sigma_y dB_{yit}, \end{aligned}$$

where the last step uses that $\ell_{it} = 0$ if $z_{it} < w_t$ by construction in equation (A.10).

Rearranging we obtain the dynamic budget constraint,

$$dn_{it} = (r_t n_{it} - c_{it} + \ell_{it}(z_{it} - w_t))dt + \ell_{it} \phi \sigma_y dB_{yit}.$$

Lemma 1. *Assume adapted processes $(c_i > 0, \ell_i \geq 0)$ satisfy (A.8)-(A.10) with $c_{i0} = \rho_e n_{i0}$. Then the plan (c_i, ℓ_i) and the associated process $n_{it} = c_{it}/\rho_e$ satisfies the dynamic budget constraint subject to initial bonds n_{i0} and the natural borrowing limit and achieves the maximum utility of the relaxed intertemporal problem. It is therefore an optimal solution of the entrepreneur's problem.*

B Worker Optimality

The worker's problem is to choose a non-negative processes (c_w, ℓ) to maximize utility $U_w(c_w, \ell)$ subject to the dynamic budget constraint

$$dn_{wt} = (n_{wt} r_t - c_{wt} + w_t \ell_t)dt,$$

with initial bonds n_{w0} given and natural borrowing limit $n_{wt} \geq -\int_t^\infty e^{-\int_t^s r_u du} w_s \ell_s ds$.

This is a standard problem. First, for any plan (c_w, ℓ) satisfying the dynamic budget constraint and natural borrowing limit we satisfy the intertemporal budget constraint

$$\int_0^\infty e^{-\int_0^t r_u du}(c_{wt} - w_t \ell_t) dt \leq n_{w0}. \tag{B.1}$$

To see this compute

$$d(e^{-\int_0^t r_u du} n_{wt}) = e^{-\int_0^t r_u du} \left(-n_{wt} r_t + n_{wt} r_t - c_{wt} + w_t \ell_t\right) dt.$$

Integrating we obtain

$$\int_t^T e^{-\int_t^s r_u du}(c_{ws} - w_s \ell_s) ds + e^{-\int_t^T r_u du} n_{wT} = n_{w0}.$$

Use the natural borrowing limit to obtain

$$\int_t^T e^{-\int_t^s r_u du} c_{ws} ds - \int_t^\infty e^{-\int_t^\infty r_u du} w_s \ell_s ds \leq n_{w0},$$

and taking the limit $T \to \infty$ we obtain the intertemporal budget constraint. We can therefore consider the relaxed problem of maximizing utility subject to the intertemporal budget constraint.

The first-order conditions of the relaxed problem are

$$e^{-\rho_w t} c_{wt}^{-1} = \lambda e^{-\int_0^t r_u du} \tag{B.2}$$

$$e^{-\rho_w t} \ell_t^{1/\psi} = \lambda e^{-\int_0^t r_u du} w_t \Longleftrightarrow \ell_t^{1/\psi} = c_{wt}^{-1} w_t, \tag{B.3}$$

and λ is set so the intertemporal budget constraint (B.1) holds with equality. These conditions are also sufficient for optimality.

From (B.2), taking time-derivatives, we obtain the Euler equation

$$dc_{wt}/c_{wt} = (r_t - \rho_w)dt. \tag{B.4}$$

A plan (c_w, ℓ) satisfying the first-order conditions (B.4) and (B.3) and the intertemporal budget constraint with equality, achieves the optimum of the relaxed problem.

It remains to show this plan can be implemented with the dynamic budget constraint. Set

$$n_{wt} = \int_t^\infty e^{-\int_t^s r_u du}(c_{ws} - w_s \ell_s)ds, \tag{B.5}$$

which satisfies the initial condition by construction. Taking derivatives with respect to time, we obtain the dynamic budget constraint

$$dn_{wt} = (n_{wt} r_t - c_{wt} + w_t \ell_t)dt.$$

Since $c_w \geq 0$, the natural borrowing limit is automatically satisfied.

Lemma 2. *Assume adapted processes* $(c_w > 0, \ell \geq 0)$ *satisfy* (B.4) *and* (B.3) *and the intertemporal budget constraint with equality.Then the plan* (c_w, ℓ) *and the associated process* n_w *given by* (B.5) *satisfies the dynamic budget constraint subject to initial bonds* n_{w0} *and the natural borrowing limit and achieves the maximum utility of the relaxed intertemporal problem. It is therefore an optimal solution of the worker's problem.*

C Proof of Proposition 1

We want to show that the prices and allocations $\mathcal{C} = (c_w, \ell, c_i, \ell_i, r, w)$ constructed as described in Section 2.3 constitute a competitive equilibrium. To do so we need to show that allocations and prices are in their permissible domain, that markets clear, and that agents are solving their problem taking prices as given.

C.1 Market Clearing

First, since $w_t, c_{it}, c_{wt} > 0$, we have that $\ell_{it} = \sigma_c(z_{it}, \eta_t)\frac{c_{it}}{\rho_e \phi \sigma_y} \geq 0$ and $\ell_t = c_{wt}^{-1} w_t > 0$. Furthermore, by construction $\int_0^1 n_{it} di + n_{wt} = 0$.

To show that markets clear we can use that the KFE is derived from the Euler equations of entrepreneurs and the worker, yielding that

$$\frac{\int_{\{z_{it}=z\}} c_{it} di}{c_{wt}} = \frac{\eta_t(z)}{\eta_{wt}}.$$

Integrating across z and using $\eta_{et} = 1 - \eta_{wt}$ and $c_{wt} = (1 - \eta_{et}) Z_t \ell_t$, we obtain

$$\int_0^1 c_{it} di = \eta_{et} \frac{c_{wt}}{1 - \eta_{et}} = \eta_{et} Z_t \ell_t,$$

and therefore

$$\int_0^1 c_{it} di + c_{wt} = Z_t \ell_t.$$

Equations (20) and (21) then ensure that market clearing (equations 5 and 6) hold.

C.2 Optimality

It remains to show that the allocations (c_i, ℓ_i) and (c_w, ℓ) satisfy the dynamic budget constraints, natural borrowing constraints, and are optimal given prices (r, w).

Since we constructed the entrepreneurs' allocations (c_i, ℓ_i) to satisfy the first-order conditions and $c_{i0} = \rho_e n_{i0}$, Lemma 1 in Appendix A ensures they satisfy budget constraints and are optimal.

The allocation satisfies the first order conditions for the worker. So given Lemma 2 in Appendix B, we only need to show that it satisfies the worker's intertemporal budget constraint with equality. We defined $n_{wt} = -\int_0^1 n_{it} di$. We therefore have that

$$
\begin{aligned}
e^{-\int_t^T r_u du} n_{wT} &= -\int_0^1 e^{-\int_t^T r_u du} n_{it} di, \\
&= \int_0^1 \left(-n_{it} + \int_t^T e^{-\int_t^s r_u du} (c_{is} - (z_{is} - w_s)\ell_{is}) ds - \int_t^T e^{-\int_t^s r_u du} \ell_{is} \phi \sigma_y dB_{iys} \right) di, \\
&= \int_0^1 -n_{it} di + \int_0^1 \int_t^T e^{-\int_t^s r_u du} (c_{is} - (z_{is} - w_s)\ell_{is}) ds di - \int_0^1 \int_t^T e^{-\int_t^s r_u du} \ell_{is} \phi \sigma_y dB_{iys} di \\
&= \left(n_{wt} + \int_0^1 \int_t^T e^{-\int_t^s r_u du} (c_{is} - (z_{is} - w_s)\ell_{is}) ds di \right).
\end{aligned}
$$

where the last term is eliminated imposing a law of large numbers.[15] We can switch the order of integration using Tonelli to obtain,

$$
e^{-\int_t^T r_u du} n_{wT} = n_{wt} + \int_t^T e^{-\int_t^s r_u du} \left(\int_0^1 c_{is} di - \int_0^1 z_{is} \ell_{is} + w_s \int_0^1 \ell_{is} di \right) ds.
$$

Use the resource constraints,

$$
e^{-\int_t^T r_u du} n_{wT} = n_{wt} + \int_t^T e^{-\int_t^s r_u du} (-c_{ws} + w_s \ell_s) ds.
$$

Use that $e^{-\int_t^T r_u du} n_{wT} = \left(\frac{e^{-\rho_w t} c_{wt}^{-1}}{c_{w0}^{-1}} \right)^{-1} e^{-\int_0^T r_u du} n_{wT}$ and take the limit $T \to \infty$. Since $\lim_{T\to\infty} e^{-\int_0^T r_u du} n_{wT} = 0$, then

$$
\int_t^\infty e^{-\int_t^s r_u du} (c_{ws} - w_s \ell_s) ds = n_{wt}, \tag{C.1}
$$

as desired.

C.3 Conclusion

Thus, initial bonds n_{w0} and n_{i0} and allocations and prices $\mathcal{C} = (c_w, \ell, c_i, \ell_i, r, w)$ that satisfy the proposition assumptions and are constructed as detailed are a *Competitive Equilibrium.*

[15] This is implied by the LLN that we imposed when writing the goods resource constraint in equation (5) because $e^{-\int_t^s r_u du}$ as a function of s is deterministic and continuous on $[0, T]$.

D Constructing the Optimal Allocation

Starting with η_0 and a path for ℓ from the Planner optimization problem, we can construct the allocation following the same steps we used to construct the competitive equilibrium using $w(\eta_t)$ and $Z(\eta_t)$. Now, however, we do not need to use the worker's first-order condition for labor supply to determine labor supply ℓ_t. Instead, we know ℓ_t and we need to find the labor tax τ_t and the initial bond distribution, n_{w0} and n_{i0}, that implement the allocation.

As in the competitive equilibrium, starting from η_0, solve the KFE forward to obtain the path of η_t, $w(\eta_t)$, and $Z(\eta_t)$. Then set the labor tax according to equation (28). We then set the representative worker's consumption to $c_{wt} = (1 - \eta_{et})Z_t\ell_t$, and use his Euler equation (10) to determine the interest rate. Set $c_{i0} = \eta_{i0}c_0$ and use entrepreneurs' Euler equation (12) and $\sigma_{cit} = \sigma_c(z_{it}, \eta_t)$ to pin down their consumption c_{it}, and set $\ell_{it} = \sigma_c(z_{it}, \eta_t)\frac{c_{it}}{\rho_e\phi\sigma_y}$. The tax rebate is $T_t = \tau_t w_t \ell_t$ and the bonds are $n_{it} = c_{it}/\rho_e$ and $n_{wt} = -\int_0^1 n_{it}di$.

Proposition 2. *Let $\mathcal{S} = (c_w, \ell, c_i, \ell_i)$ be an allocation built as described above, with associated prices (r, w), taxes (τ, T) and initial bonds (n_{i0}, n_{w0}). Assume that τ is feasible, that $w_t(1 - \tau_t), c_{it}, c_{wt} > 0$ for all t and $\lim_{T\to\infty} e^{-\int_0^T r_u du} n_{wT} = 0$. Then $\mathcal{S}$ is an optimal allocation.*

Proof. The allocation maximizes the social planner's transformed objective (31). To show it is an optimal allocation we only need to show that it is implementable. An argument analogous to the one in Proposition 1 shows that $(c_w, \ell, c_i, \ell_i, r, w, \tau, T, n_w, n_i)$ is a competitive equilibrium with tax and lump-sum transfers, so $\mathcal{S}$ is implementable. Since $w_t(1-\tau_t), c_{it}, c_{wt} > 0$, we have $\ell_{it} \geq 0$ and $\ell_t > 0$, and the first order conditions for the workers and entrepreneurs are satisfied and $n_{wt} + \int_0^1 n_{it}di = 0$ by construction. The same argument as in the proof of Proposition 1 shows that the resource constraints hold and that the dynamic budget constraint of entrepreneurs hold with the natural borrowing limit and the intertemporal budget constraints hold with equality, which imply that the plan (c_i, ℓ_i) is optimal for entrepreneur i. Since $T_t = \tau_t w_t \ell_t$, the budget constraint of the worker is the same as in the case without taxes, so the same argument as in the proof of Proposition 1 shows that the workers dynamic budget constraint and natural borrowing limit hold, and the intertemporal budget constraint holds with equality, which implies the plan (c_w, ℓ) is optimal for the worker. We conclude that $(c_w, \ell, c_i, \ell_i, r, w, \tau, T, n_w, n_i)$ is a competitive equilibrium with tax and lump-sum transfers, so $\mathcal{S}$ is implementable. $\square$

D.1 Transforming the Planner's Objective

We want to write the last term of the objective function in a way that is more useful for numerical computation,

$$\int_0^\infty \Gamma_{et} \left(\int_0^1 \log \eta_{it} di - \log \eta_{et} \right) dt.$$

Use the Euler equations to write

$$\begin{aligned} c_{wt} &= c_{w0} \exp \left(\int_0^t r - \rho_w dt \right), \\ c_{it} &= c_{i0} \exp \left(\int_0^t r - \rho_e + \sigma_{cit}^2 - \frac{1}{2} \sigma_{cit}^2 dt + \int_0^t \sigma_{cit} dB_{yit} \right). \end{aligned}$$

Putting them together, we can express the consumption of entrepreneurs in terms of workers' consumption,

$$c_{it} = c_{i0} \times \frac{c_{wt}}{c_{w0}} \times \exp \left(\int_0^t \rho_w - \rho_e + \frac{1}{2} \sigma_{cit}^2 dt + \int_0^t \sigma_{cit} dB_{yit} \right).$$

Divide and multiply by c_0 on the rights-hand-side, and then divide by c_t on both sides,

$$\frac{c_{it}}{c_t} = \frac{c_{i0}}{c_0} \times \frac{c_{wt}}{c_t} \times \frac{c_0}{c_{w0}} \times \exp \left(\int_0^t \rho_w - \rho_e + \frac{1}{2} \sigma_{cit}^2 dt + \int_0^t \sigma_{cit} dB_{yit} \right).$$

Now take logs,

$$\log \eta_{it} = \log \eta_{wt} + (\log \eta_{i0} - \log \eta_{w0}) + \int_0^t \rho_w - \rho_e + \frac{1}{2} \sigma_{cit}^2 dt + \int_0^t \sigma_{cit} dB_{yit}.$$

Integrate over i, and use a LLN and the fact that σ_{cit} is bounded because τ is feasible,

$$\int_0^1 \log \eta_{it} di = \log \eta_{wt} + \left(\int_0^1 \log \eta_{i0} di - \log \eta_{w0} \right) + \int_0^t \left(\rho_w - \rho_e + \int_0^1 \frac{1}{2} \sigma_{cit}^2 di \right) dt.$$

Now subtract $\log \eta_{et}$, multiply by Γ_{et}, and integrate over time,

$$\begin{aligned}
&\int_0^\infty \Gamma_{et} \left(\int_0^1 \log \eta_{it} di - \log \eta_{et} \right) dt = \\
&\quad \int_0^\infty \Gamma_{et} \left(\log \eta_{wt} - \log \eta_{et} + \left(\int_0^1 \log \eta_{i0} di - \log \eta_{w0} \right) + \int_0^t \left(\rho_w - \rho_e + \int_0^1 \frac{1}{2} \sigma_{cis}^2 di \right) ds \right) dt.
\end{aligned}$$

Finally, change the order of integration for the last term

$$\int_0^\infty \Gamma_{et} \int_0^t \left(\rho_w - \rho_e + \int_0^1 \frac{1}{2} \sigma_{cis}^2 di \right) ds dt = \int_0^\infty \Gamma_{et} \frac{1}{\rho_e} \left(\rho_w - \rho_e + \frac{1}{2} \int_0^1 \sigma_{cit}^2 di \right) dt.$$

We obtain the final expression,

$$\begin{aligned}
&\int_0^\infty \Gamma_{et} \left(\int_0^1 \log \eta_{it} di - \log \eta_{et} \right) dt = \\
&\quad \int_0^\infty \Gamma_{et} \left(\log \eta_{wt} - \log \eta_{et} + \left(\int_0^1 \log \eta_{i0} di - \log \eta_{w0} \right) + \frac{1}{\rho_e} \left(\rho_w - \rho_e + \frac{1}{2} \int_0^1 \sigma_{cit}^2 di \right) \right) dt.
\end{aligned} \tag{D.1}$$

Special Case of $z_{it} = \bar{z}$. In the special case with only one type, $z_{it} = \bar{z}$, the expression simplifies. Start with

$$c_{it} = c_{i0} \exp \left(\int_0^t r - \rho_e + \sigma_{ct}^2 - \frac{1}{2} \sigma_{ct}^2 dt + \int_0^t \sigma_{ct} dB_{yit} \right).$$

Integrate across i using the LLN and σ_{ct} bounded, to obtain

$$\int_0^1 c_{it} di = \int_0^1 c_{i0} di \times \exp \left(\int_0^t r - \rho_e + \sigma_{ct}^2 \right).$$

Plug back into the expression for c_{it},

$$c_{it} = \frac{c_{i0}}{\int_0^1 c_{i0} di} \int_0^1 c_{it} di \exp\left(\int_0^t -\frac{1}{2}\sigma_{ct}^2 dt + \int_0^t \sigma_{ct} dB_{yit} \right).$$

Divide by c_t, and use $\eta_{it} = c_{it}/c_t$ and $\eta_{et} = \int_0^1 c_{it} di/c_t$,

$$\eta_{it} = \frac{c_{i0}}{\int_0^1 c_{i0} di} \eta_{et} \exp\left(\int_0^t -\frac{1}{2}\sigma_{ct}^2 dt + \int_0^t \sigma_{ct} dB_{yit} \right).$$

Given that we have equal weights for entrepreneurs and they all have the same productivity, we can without loss of generality take $c_{i0} = \int_0^1 c_{i0} di$. Take logs and integrate across i, again using a LLN to eliminate the last term,

$$\int_0^1 \log \eta_{it} di - \log \eta_{et} = \int_0^t -\frac{1}{2}\sigma_{ct}^2 dt.$$

Now multiply by Γ_{et} and integrate over t, and switch the order of integration to obtain

$$\int_0^\infty \Gamma_{et} \left(\int_0^1 \log \eta_{it} di - \log \eta_{et} \right) dt = -\int_0^\infty \Gamma_{et} \frac{1}{\rho_e} \frac{1}{2} \sigma_{ct}^2 dt. \tag{D.2}$$

E Case with $\rho_e = \rho_w$

The main body of the paper assumes $\rho_e > \rho_w$. This is done to obtain a non-degenerate stationary wealth distribution in the competitive equilibrium, avoiding entrepreneurs' precautionary saving leading them to accumulate all the wealth. But our results are robust to the case with $\rho_e = \rho_w$. In this case, entrepreneurs' precautionary saving motive σ_{cit}^2—which is always bounded away from zero for the positive measure of entrepreneurs who produce—makes their expected consumption growth larger than workers' consumption growth (compare Euler equations (10) and (12)).[16] This means when $\rho_e = \rho_w$, in the long run, entrepreneurs

[16] The left hand side of equation (20) is bounded above by $\frac{1}{\rho_e(\phi\sigma_y)^2}\bar{z}(\bar{z} - w)$, which means w is bounded away below $\bar{z}$. Equation (22) shows σ_{cit} is thus bounded away from zero for all $z_{it} > w_t$.

accumulate all the wealth, i.e., $\eta_{et} \to 1$.

The expressions for finite t in the main body of the paper are still valid. The only results that are affected are those pertaining to the long run. Although their precise statement must be adjusted, the spirit remains the same.

Entrepreneur's Consumption Share is Greater than their Pareto Weight in the Long Run. In the baseline case with $\rho_e > \rho_w$ Pareto weights $\frac{\Gamma_{et}}{\Gamma_{et}+\Gamma_{wt}} \to 0$ while the entrepreneurs' share of consumption does not because of their precautionary saving. For example, in the special case with only one productivity type,

$$\eta_{et} \to \eta_e^* = \min\left\{ \frac{1}{\sqrt{\rho_e - \rho_w}} \frac{\rho_e \phi \sigma_y}{\bar{z}}, 1 \right\}, \tag{E.1}$$

When $\rho_e = \rho_w$, however, entrepreneurs' Pareto weights are constant and less than 1, $\frac{\Gamma_{et}}{\Gamma_{et}+\Gamma_{wt}} = \frac{\Gamma_{e0}}{\Gamma_{e0}+\Gamma_{w0}} \in (0, 1)$. Just as in the $\rho_e > \rho_w$ case, entrepreneurs precautionary saving makes $\eta_{et} \to 1$, larger than their Pareto weight in the long run.

In the Long Run the Optimal Labor Tax is Positive (Not a Subsidy). In Section 3.4 we derive equation (45) for the optimal tax in the long run, under the assumption that η_t converges to a non-degenerate stationary distribution. A key result is that, in the long run, the planner uses a tax on labor,

$$1 - \tau^* = \frac{c_w^*}{w^* \ell^*} < 1,$$

in contrast to the typical optimal subsidy in the monopolistic competition markup literature. This result is still true if $\rho_e = \rho_w$, and in a more extreme form. First, the general expression for the optimal tax (not just in the long run) is

$$1 - \tau_t = (1 + \pi_t) \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{et}+\Gamma_{wt}}}.$$

Use $1+\pi_t = Z_t/w_t$ and $\eta_{wt} = c_{wt}/(Z_t \ell_t)$ to write

$$1-\tau_t = \frac{c_{wt}}{w_t \ell_t} \frac{1}{\frac{\Gamma_{wt}}{\Gamma_{et}+\Gamma_{wt}}}. \tag{E.2}$$

When $\rho_e > \rho_w$, the term $\frac{\Gamma_{wt}}{\Gamma_{et}+\Gamma_{wt}} \to 1$. If $c_{wt}/(w_t\ell_t)$ converges, it must be to something positive but less than one. In the long run, workers have positive debt so their consumption must be less than their labor income (the tax on labor and the lump-sum transfers cancel out), but they are not up against their natural borrowing limit. From this we get get the result that $\tau_t \to \tau^* > 0$.

When $\rho_e = \rho_w$, the term $\frac{\Gamma_{wt}}{\Gamma_{et}+\Gamma_{wt}} = \frac{\Gamma_{w0}}{\Gamma_{e0}+\Gamma_{w0}} \in (0,1)$ for all t. The planner always sets labor to the first-best level and aggregate TFP converges to $Z(\eta^*)$, so output converges to $c^* > 0$. Since $\eta_{et} \to 1$, then $n_{et} \to n_{e^*} = \frac{c^*}{\rho_e} > 0$ and $c_{wt} \to 0$. By bond market clearing, $n_w^* = -n_e^* < 0$, so workers natural borrowing limit must be strictly positive. If $w_t \to 0$, then the natural borrowing constraint would be zero (because ℓ^{FB} is finite), which cannot be. Therefore $c_{wt}/(w_t\ell_t) \to 0$. This yields $\tau_t \to 1$. In the long run, the optimal labor tax is not only positive (a tax), but it approaches 100%. As workers become immiserated, the income effect raises labor supply, and an extreme tax is necessary to implement the finite first-best employment.

In the Long Run the Optimal Tax on Entrepreneurs' Saving Eliminates their Precautionary Saving and Reduces their Consumption Share. In Section 3.5.1 we introduce a tax on entrepreneurs' savings. In the baseline with $\rho_e > \rho_w$ we have $\eta_{et} \to 0$, while without the tax we have (E.1). We obtain

$$\lim_{t\to\infty} \frac{1-\tau_t}{1+\pi_t} = 1,$$
$$\lim_{t\to\infty} \frac{\tau_{st}}{\sigma_{ct}^2} = 1.$$

In the long run, the planner uses the labor tax to exactly eliminate the markup, and the savings tax to exactly eliminate precautionary saving. If instead $\rho_e = \rho_w$, η_{et} is constant in the optimal allocation with a savings tax,

$$\eta_{et} = \eta_e = \frac{\Gamma_{e0}}{\Gamma_{e0} + \Gamma_{w0}} \times \left(1 + (1 - \eta_e)\frac{\sigma_c^2(\eta_e)}{\rho_e}\right) > \frac{\Gamma_{e0}}{\Gamma_{e0} + \Gamma_{w0}}, \tag{E.3}$$

while without the tax $\eta_{et} \to 1$, as discussed above. In both cases, the savings tax is used to keep entrepreneurs' long-run share of consumption below what it would be without the tax. The optimal savings tax satisfies $\tau_{st} = \sigma_{ct}^2$ at all time (not just in the long-run limit). The main difference with the $\rho_e > \rho_w$ case is that

$$\lim_{t\to\infty} \frac{1 - \tau_t}{1 + \pi_t} = \lim_{t\to\infty} \frac{\eta_{wt}}{\frac{\Gamma_{wt}}{\Gamma_{et} + \Gamma_{wt}}} < 1. \tag{E.4}$$

The planner uses the savings tax to eliminate the entrepreneurs' precautionary saving and keep their consumption share constant. But he keeps it at a level above their Pareto weight in order to reduce their exposure to risk. To undo the effect of this deviation from Pareto weights on labor supply, he needs to raise the labor tax relative to the subsidy that fully eliminates the markup.

F Derivation of KFE for $\eta_t(z)$

F.1 Derive process for η_{it}

Keep track of (z_i, η_i) where $\eta_{it} = c_{it}/c_t$. We need to compute the growth rate of aggregate consumption c_t:

$$c_{wt}\mu_{cwt} + \int c_{it}\mu_{cit} di = \mu_{ct} c_t$$

$$\eta_{wt}\mu_{cwt} + \int \eta_t(z)\mu_{ct}(z) dz = \mu_{ct}$$

(using the fact that growth rates depend only on z type).

Now use the Euler equations to rewrite this as

$$\eta_{wt}(r_t - \rho_w) + \int \eta_t(z) \left(r_t - \rho_e + \frac{(z_t - w_t)^2}{\phi^2 \sigma_y^2} \right) dz = \mu_{ct}$$

$$\left(\underbrace{\eta_{wt} + \int \eta_t(z) dz}_{=1} \right) r_t - \underbrace{\left(\eta_{wt}\rho_w + \rho_e \int \eta_t(z) dz \right)}_{=\bar{\rho}_t} + \int \eta_t(z) \frac{(z_t - w_t)^2}{\phi^2 \sigma_y^2} dz = \mu_{ct}$$

$$\mu_{ct} = r_t - \bar{\rho}_t + \int \eta_t(z) \frac{(z_t - w_t)^2}{\phi^2 \sigma_y^2} dz.$$

So now we can compute the law of motion of $\eta_{it} = c_{it}/c_t$:

$$d(c_{it}/c_t) = \left(\frac{c_{it}}{c_t} \right) (\mu_{cit} - \mu_{ct})\, dt + \left(\frac{c_{it}}{c_t} \right) \sigma_{cit} dB_{it}$$

$$d(c_{it}/c_t) = \left(\frac{c_{it}}{c_t} \right) \underbrace{\left(\bar{\rho}_t - \rho_e + \frac{(z_t - w_t)^2}{\phi^2 \sigma_y^2} - \int \eta_t(z) \frac{(z_t - w_t)^2}{\phi^2 \sigma_y^2} dz \right)}_{\mu_\eta(z,t)} dt$$

$$+ \left(\frac{c_{it}}{c_t} \right) \underbrace{\frac{(z_t - w_t)^+}{\phi \sigma_y}}_{\sigma_{ct}(z)} dB_{yt},$$

or, more compactly,

$$d\eta_{it} = \eta_{it}\mu_\eta(z,t)dt + \eta_{it}\sigma_{ct}(z)dB_{yt}.$$

F.2 Derive KFE

Let $D(\eta, z, t)$ be the joint distribution of η, z. $D(\eta, z, t)$ is the measure of entrepreneurs with $z_i = z$ and $\eta_i = \eta$. This distribution has a convenient mapping into $\eta(z,t)$ as

$$\eta(z,t) = \int D(\eta, z, t)\eta d\eta$$

Let $X = (z_i, \eta_i)$

$$dX_t = \begin{bmatrix} \mu_z(z) \\ \mu_\eta(z,t)\eta \end{bmatrix} dt + \begin{bmatrix} \sigma_z(z) & 0 \\ 0 & \sigma_c(z,\eta_t)\eta \end{bmatrix} \begin{bmatrix} dB_{zt} \\ dB_{yt} \end{bmatrix}.$$

In order to derive a KFE for the joint distribution $D(\cdot,\cdot)$, we re-write dX_t as a function of a standard multivariate Wiener process with independent component. Consider two independent Wiener processes W_{1t}, W_{2t}:

$$\begin{aligned} B_{zt} &= W_{1t} \\ B_{yt} &= \lambda_{yz} W_{1t} + \sqrt{1-\lambda_{yz}^2}\, W_{2t}. \end{aligned}$$

Therefore, we have

$$dX_t = \begin{bmatrix} \mu_z(z) \\ \mu_\eta(z,t)\eta \end{bmatrix} dt + \begin{bmatrix} \sigma_z(z) & 0 \\ \sigma_c(z,\eta_t)\eta\lambda_{yz} & \sigma_c(z,\eta_t)\eta\sqrt{1-\lambda_{yz}^2} \end{bmatrix} \begin{bmatrix} dW_{1t} \\ dW_{2t} \end{bmatrix}.$$

To simplify the notation, we can write more generally

$$dX_t = \begin{bmatrix} \mu_z(z) \\ \mu_\eta(z,t)\eta \end{bmatrix} dt + \begin{bmatrix} \sigma_{z1}(z) & \sigma_{z2}(z) \\ \sigma_{c1}(z,t)\eta & \sigma_{c2}(z,t)\eta \end{bmatrix} \begin{bmatrix} dW_{1t} \\ dW_{2t} \end{bmatrix}.$$

We derive the KFE for $D(\eta, z, t)$, starting from the first order terms[17]

$$\begin{aligned} \partial D(\eta,z,t) =& -\frac{\partial}{\partial z}\left[\mu_z(z)D(\eta,z,t)\right] - \frac{\partial}{\partial \eta}\left[\mu_\eta(z,t)\eta D(\eta,z,t)\right] + \text{ 2nd order terms} \\ =& -\left[\mu_z'(z)D(\eta,z,t) + \mu_z(z)D_z'(\eta,z,t)\right] \\ & - \left[\mu_\eta(z)D(\eta,z,t) + \mu_\eta(z)\eta D_\eta'(\eta,z,t)\right] + \text{ 2nd order terms.} \end{aligned}$$

[17]Since $d\eta_{it}/\eta_{it} = dc_{it}/c_{it}$, we have $d\eta_i = \mu_c\eta_i dt + \sigma_c\eta_i dB_t$.

Now, we define the second order terms. The variance covariance matrix is

$$\Lambda = \begin{bmatrix} \sigma_{z1}(z) & \sigma_{z2}(z) \\ \sigma_{c1}(z,t)\eta & \sigma_{c2}(z,t)\eta \end{bmatrix} \begin{bmatrix} \sigma_{z1}(z) & \sigma_{z2}(z) \\ \sigma_{c1}(z,t)\eta & \sigma_{c2}(z,t)\eta \end{bmatrix}'$$

$$\Lambda = \begin{bmatrix} \sigma_{z1}^2(z) + \sigma_{z2}^2(z) & \eta\sigma_{z1}(z)\sigma_{c1}(z) + \eta\sigma_{z2}(z)\sigma_{c2}(z) \\ \eta\sigma_{z1}(z)\sigma_{c1}(z) + \eta\sigma_{z2}(z)\sigma_{c2}(z) & \sigma_{c1}^2(z)\eta^2 + \sigma_{c2}^2(z)\eta^2 \end{bmatrix},$$

$$\text{2nd order terms} = \frac{1}{2}\frac{\partial^2}{\partial z^2}[\Lambda_{11}D(\eta,z,t)] + \frac{1}{2}\frac{\partial^2}{\partial \eta^2}[\Lambda_{22}D(\eta,z,t)] + \frac{\partial^2}{\partial \eta z}[\Lambda_{12}D(\eta,z,t)].$$

Further evaluating the matrix by elements, the first term is:

$$\begin{aligned}
\frac{1}{2}\frac{\partial^2}{\partial z^2}[\Lambda_{11}D(\eta,z,t)] &= \frac{1}{2}\frac{\partial^2}{\partial z^2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))D(\eta,z,t)] \\
&= \frac{1}{2}\frac{\partial}{\partial z}[2(\sigma_{z1}(z)\sigma_{z1}'(z) + \sigma_{z2}(z)\sigma_{z2}'(z))D(\eta,z,t) + (\sigma_{z1}^2(z) + \sigma_{z2}^2(z))D_z(\eta,z,t)] \\
&= \frac{\partial}{\partial z}[(\sigma_{z1}(z)\sigma_{z1}'(z) + \sigma_{z2}(z)\sigma_{z2}'(z))D(\eta,z,t)] \\
&\quad + \frac{1}{2}\frac{\partial}{\partial z}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))D_z(\eta,z,t)] \\
&= [\left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right) D(\eta,z,t)] + [\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right) D_z(\eta,z,t)] \\
&\quad + [\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right) D_z(\eta,z,t)] + \frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))D_{zz}(\eta,z,t)] \\
&= [\left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right) D(\eta,z,t)] + 2[\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right) D_z(\eta,z,t)] \\
&\quad + \frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))D_{zz}(\eta,z,t)].
\end{aligned}$$

The second term is:

$$
\begin{aligned}
\frac{1}{2}\frac{\partial^2}{\partial\eta^2}[\Lambda_{22}D(\eta,z,t)] =&\frac{1}{2}\frac{\partial^2}{\partial\eta^2}[(\sigma_{c1}^2(z)\eta^2+\sigma_{c2}^2(z)\eta^2)D(\eta,z,t)] \\
=&(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\frac{1}{2}\frac{\partial^2}{\partial\eta^2}[\eta^2 D(\eta,z,t)] \\
=&(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\frac{1}{2}\frac{\partial}{\partial\eta}[2\eta D(\eta,z,t)+\eta^2 D_\eta(\eta,z,t)] \\
=&(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\left[\frac{\partial}{\partial\eta}[\eta D(\eta,z,t)]+\frac{1}{2}\frac{\partial}{\partial\eta}[\eta^2 D_\eta(\eta,z,t)]\right] \\
=&(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\left[[D(\eta,z,t)+\eta D_\eta(\eta,z,t)]+\eta D_\eta(\eta,z,t)+\frac{1}{2}[\eta^2 D_{\eta\eta}(\eta,z,t)]\right] \\
=&(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\left[[D(\eta,z,t)+2\eta D_\eta(\eta,z,t)]+\frac{1}{2}[\eta^2 D_{\eta\eta}(\eta,z,t)]\right].
\end{aligned}
$$

The third term is:

$$
\begin{aligned}
\frac{\partial^2}{\partial\eta\partial z}[\Lambda_{12}D(\eta,z,t)] &= \frac{\partial^2}{\partial\eta\partial z}[[\eta\sigma_{z1}(z)\sigma_{c1}(z)+\eta\sigma_{z2}(z)\sigma_{c2}(z)]D(\eta,z,t)] \\
&= \frac{\partial^2}{\partial\eta\partial z}[\sum_{j=1}^{2}(\sigma_{zj}(z)\sigma_{cj}(z))\eta D(\eta,z,t)] \\
&= \frac{\partial}{\partial z}[\sum_{j=1}^{2}(\sigma_{zj}(z)\sigma_{cj}(z))D(\eta,z,t)]+\frac{\partial}{\partial z}[\sum_{j=1}^{2}(\sigma_{zj}(z)\sigma_{cj}(z))\eta D_\eta(\eta,z,t)] \\
&= \sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z)+\sigma_{zj}(z)\sigma_{cj}'(z)\right]D(\eta,z,t)+\sum_{j=1}^{2}[\sigma_{zj}(z)\sigma_{cj}(z)D_z(\eta,z,t)] \\
&+\sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z)+\sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta D_\eta(\eta,z,t) \\
&+\sum_{j=1}^{2}[\sigma_{zj}(z)\sigma_{cj}(z))\eta D_{\eta z}(\eta,z,t)].
\end{aligned}
$$

Combining all first and second order terms yields

$$
\begin{aligned}
\partial D(\eta,z,t) =& -\left[\mu_z'(z)D(\eta,z,t)+\mu_z(z)D_z'(\eta,z,t)\right]-\left[\mu_\eta(z)D(\eta,z,t)+\mu_\eta(z)\eta D_\eta'(\eta,z,t)\right]\\
&\qquad +\left(\sum_j \sigma_{zj}'(z)^2+\sigma_{zj}(z)\sigma_{zj}''(z)\right)D(\eta,z,t)+2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right)D_z(\eta,z,t)\\
&\qquad\quad +\frac{1}{2}[(\sigma_{z1}^2(z)+\sigma_{z2}^2(z))D_{zz}(\eta,z,t)]\\
&+(\sigma_{c1}^2(z)+\sigma_{c2}^2(z))\left[[D(\eta,z,t)+2\eta D_\eta(\eta,z,t)]+\frac{1}{2}[\eta^2 D_{\eta\eta}(\eta,z,t)]\right]\\
&+\sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z)+\sigma_{zj}(z)\sigma_{cj}'(z)\right]D(\eta,z,t)+\sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z)D_z(\eta,z,t)\right]\\
&\quad +\sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z)+\sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta D_\eta(\eta,z,t)+\sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z))\eta D_{\eta z}(\eta,z,t)\right].
\end{aligned}
$$

To obtain $\eta(z,t)$, use that

$$
\begin{aligned}
\eta(z,t) &= \int \eta D(\eta,z,t)d\eta\\
\partial_t\eta(z,t) &= \int \eta\partial D(\eta,z,t)d\eta\\
\eta_z(z,t) &= \int \eta D_z(\eta,z,t)d\eta\\
\eta_{zz}(z,t) &= \int \eta D_{zz}(\eta,z,t)d\eta.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\partial\eta(z,t) = &- \left[\mu_z'(z)\eta(z,t) + \mu_z(z)\eta_z'(\eta,z,t)\right] - \left[\mu_\eta(z)\eta(z,t) + \mu_\eta(z)\int \eta^2 D_\eta'(\eta,z,t)d\eta\right] \\
&+ \left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right)\eta(z,t) + 2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right)\eta_z'(z,t) \\
&+ \frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))\eta_{zz}''(\eta,z,t)] \\
&+ (\sigma_{c1}^2(z) + \sigma_{c2}^2(z))\left[[\eta(z,t) + 2\int \eta^2 D_\eta(\eta,z,t)d\eta] + \frac{1}{2}[\int \eta^3 D_{\eta\eta}(\eta,z,t)d\eta]\right] \\
&+ \sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta(z,t) + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z)\eta_z'(z,t)\right] \\
&+ \sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\int \eta^2 D_\eta(\eta,z,t)d\eta + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z))\int \eta^2 D_{\eta z}(\eta,z,t)d\eta\right].
\end{aligned}
$$

Conjecture that $\eta^2 D(z,\eta,t)\big|_0^\infty = \lim_{\eta\to\infty}\eta^2 D(z,\eta,t) - \lim_{\eta\to 0}\eta^2 D(z,\eta,t) = 0-0 = 0$. Then, using integration by parts,

$$
\int \eta^2 D_\eta(z,\eta,t)d\eta = \eta^2 D(z,\eta,t)\big|_0^\infty - 2\int_0^\infty \eta D(z,\eta,t)d\eta = -2\eta(z,t).
$$

Substitute this into the KFE to obtain

$$
\begin{aligned}
\partial\eta(z,t) = &- \left[\mu_z'(z)\eta(z,t) + \mu_z(z)\eta_z'(\eta,z,t)\right] - \left[\mu_\eta(z)\eta(z,t) + \mu_\eta(z)(-2\eta(z,t))\right] \\
&+ \left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right)\eta(z,t) + 2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right)\eta_z'(z,t) \\
&+ \frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))\eta_{zz}''(\eta,z,t)] \\
&+ (\sigma_{c1}^2(z) + \sigma_{c2}^2(z))\left[[\eta(z,t) + 2(-2\eta(z,t))] + \frac{1}{2}[\int \eta^3 D_{\eta\eta}(\eta,z,t)d\eta]\right] \\
&+ \sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta(z,t) + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z)\eta_z'(z,t)\right] \\
&+ \sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right](-2\eta(z,t)) + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z))\int \eta^2 D_{\eta z}(\eta,z,t)d\eta\right].
\end{aligned}
$$

Conjecture that $\eta^3 D_\eta(z,\eta,t)\big|_0^\infty = \lim_{\eta\to\infty} \eta^3 D_\eta(z,\eta,t) - \lim_{\eta\to 0} \eta^3 D_\eta(z,\eta,t) = 0 - 0 = 0$. Then, using integration by parts,

$$\int \eta^3 D_{\eta\eta}(\eta,z,t)d\eta = \eta^3 D_\eta(z,\eta,t)\big|_0^\infty - 3\int \eta^2 D_\eta(\eta,z,t)d\eta = 6\eta(z,t).$$

Again, substitute this into the KFE to obtain

$$\begin{aligned}
\partial\eta(z,t) =& -\left[\mu_z'(z)\eta(z,t) + \mu_z(z)\eta_z'(\eta,z,t)\right] - \left[\mu_\eta(z)\eta(z,t) + \mu_\eta(z)(-2\eta(z,t))\right] \\
&+ \left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right)\eta(z,t) + 2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right)\eta_z'(z,t) \\
&+ \frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))\eta_{zz}''(\eta,z,t)] \\
&+ (\sigma_{c1}^2(z) + \sigma_{c2}^2(z))\left[[\eta(z,t) + 2(-2\eta(z,t))] + \frac{1}{2}[6\eta(z,t)]\right] \\
&+ \sum_{j=1}^2 \left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta(z,t) + \sum_{j=1}^2 \left[\sigma_{zj}(z)\sigma_{cj}(z)\eta_z'(z,t)\right] \\
&+ \sum_{j=1}^2 \left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right](-2\eta(z,t)) + \sum_{j=1}^2 \left[\sigma_{zj}(z)\sigma_{cj}(z))\int \eta^2 D_{\eta z}(\eta,z,t)d\eta\right].
\end{aligned}$$

A similar argument can be applied to obtain

$$\int \eta^2 D_{\eta z}(\eta,z,t)d\eta = -2\eta_z'(z,t),$$

which implies that

$$
\begin{aligned}
\partial\eta(z,t) =& - \left[\mu_z'(z)\eta(z,t) + \mu_z(z)\eta_z'(\eta,z,t)\right] - \left[\mu_\eta(z)\eta(z,t) + \mu_\eta(z)(-2\eta(z,t))\right] \\
&+ \left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right)\eta(z,t) + 2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right)\eta_z'(z,t) \\
&+\frac{1}{2}[(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))\eta_{zz}''(\eta,z,t)] \\
&+(\sigma_{c1}^2(z) + \sigma_{c2}^2(z))\left[[\eta(z,t) + 2(-2\eta(z,t))] + \frac{1}{2}[6\eta(z,t)]\right] \\
&+\sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\eta(z,t) + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z)\eta_z'(z,t)\right] \\
&+\sum_{j=1}^{2}\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right](-2\eta(z,t)) + \sum_{j=1}^{2}\left[\sigma_{zj}(z)\sigma_{cj}(z))(-2\eta_z'(z,t))\right].
\end{aligned}
$$

Grouping terms yields

$$
\begin{aligned}
\partial\eta(z,t) =& \eta(z,t)\left[-\mu_z'(z) + \mu_\eta(z) + \left(\sum_j \sigma_{zj}'(z)^2 + \sigma_{zj}(z)\sigma_{zj}''(z)\right) - \sum_j\left[\sigma_{zj}'(z)\sigma_{cj}(z) + \sigma_{zj}(z)\sigma_{cj}'(z)\right]\right] \\
&+ \eta_z'(z,t)\left[-\mu_z(z) + 2\left(\sum_j \sigma_{zj}(z)\sigma_{zj}'(z)\right) - \sum_j \sigma_{zj}(z)\sigma_{cj}(z)\right] \\
&+ \eta_{zz}''(\eta,z,t)\left[\frac{1}{2}(\sigma_{z1}^2(z) + \sigma_{z2}^2(z))\right].
\end{aligned}
$$

In our model

$$
\begin{aligned}
&\sigma_{z1}(z) = \sigma_z(z) \\
&\sigma_{z2}(z) = 0 \\
&\sigma_{c1}(z) = \sigma_c(z,\eta_t)\lambda_{yz} \\
&\sigma_{c2}(z) = \sigma_c(z,\eta_t)\sqrt{1-\lambda_{yz}^2}.
\end{aligned}
$$

Therefore, the KFE is

$$
\begin{aligned}
\partial\eta(z,t) = \eta(z,t) & \left[-\mu_z'(z) + \mu_\eta(z,t) + \sigma_z'(z)^2 + \sigma_z(z)\sigma_z''(z) - \left(\sigma_z'(z)\sigma_c(z,\eta_t) + \sigma_z(z)\sigma_c'(z,t)\right)\lambda_{yz}\right] \\
& + \eta_z'(z,t)\left[-\mu_z(z) + 2\sigma_z(z)\sigma_z'(z) - \sigma_z(z)\sigma_c(z,\eta_t)\lambda_{yz}\right] \\
& + \eta_{zz}''(\eta,z,t)\left[\frac{1}{2}\sigma_z^2(z)\right]
\end{aligned}
$$

where $\mu_z(z), \sigma_z(z), \lambda_{yz}$ are primitives, and $\sigma_c(z,\eta_t), \mu_\eta(z,t)$ are equilibrium objects given by:

$$
\begin{aligned}
\sigma_c(z,\eta_t) &= \frac{(z_t - w_t)^+}{\phi\sigma_y} \\
\mu_\eta(z,t) &= \bar{\rho}_t - \rho_e + \frac{(z_t - w_t)^2}{\phi^2\sigma_y^2} - \int \eta_{zt}\frac{(z_t - w_t)^2}{\phi^2\sigma_y^2}dz.
\end{aligned}
$$

G Numerical Example Details

Parameter values are reported in Table G.1. We set the workers discount factor ρ equal to 0.035, and entrepreneurs discount factor ρ^e equal to 0.097 as in Di Tella and Hall (2022).[18] We set the Frisch elasticity ψ equal to one, a conventional value in the literature that lies between micro and macro estimates Chetty, Guren, Manoli, and Weber (2011).

Finally, we specify parameters for the stochastic process driving fluctuations in firms' output and productivity. The law of motions of entrepreneurs' net worth and productivity are characterized by

$$
\begin{aligned}
dn_{it} &= \left(r_t n_{it} - c_{it} + (z_{it} - w_t)\,\ell_{it}\right)dt + (\phi\sigma_y)\,\ell_{it}dB_{yit} \\
dz_{it} &= \mu_z\left(z_{it}\right)dt + \sigma_z\left(z_{it}\right)dB_{zit}
\end{aligned}
$$

where B_{yit} and B_{zit} are two Brownian motions specific to the entrepreneur that have correlation $\lambda_{yz}, \sigma_y > 0$, and $\sigma_z\left(z_{it}\right) > 0$.

[18] These parameter values imply that in steady state the annual interest rate is 3.5% and the consumption share of entrepreneurs is 20%.

Parameter	Description	Value
ψ	Frisch Elasticity	1
ρ	Household's discount factor	0.035
ρ_e	Entrepreneurs' discount factor	0.097
θ	Speed of mean reversion of z_{it}	0.3
σ_y	Volatility of innovations to output, B_{yt}	0.1
$\sigma_{\log z}$	Volatility of innovations to $\log(z)$, B_{zt}	0.1
λ_{yz}	Correlation of B_{yt}, B_{zt}	1
ϕ	Exposure to idiosyncratic risk	1

Table G.1: Parameter values for the numerical illustration.

In order to solve the model numerically, we impose more structure on the stochastic process for firms' productivity. We assume that $\log z$ follows a standard Ornstein-Uhlenbeck process according to equation (G.1).

$$d\log z_{it} = \theta\left(\log \bar{z} - \log z_{it}\right) dt + \sigma_{\log z} dB_{zt} \tag{G.1}$$

Moreover, we set $\log(\bar{z}) = -\frac{1}{\theta}\frac{1}{2}\sigma^2_{\log z}$, so that the implied stochastic process for z takes a simple form:

$$\begin{aligned} dz_{it} &= z_{it}\left(-\theta\log\left(z_{it}\right) + \theta\log(\bar{z}) + \frac{1}{2}\sigma^2_{\log z}\right) dt + z_{it}\sigma_{\log z} dB_{zt} \\ &= -\theta z_{it}\log\left(z_{it}\right) dt + z_{it}\sigma_{\log z} dB_{zt} \end{aligned}$$

Therefore, the stochastic process driving fluctuations in firms' output and productivity is disciplined by a set of parameters $(\sigma_y, \sigma_{\log z}, \lambda_{yz}, \theta)$.

We set $\theta = 0.3$ and $\sigma_{logz} = 0.1$, meaning that the annual autocorrelation of firm productivity is equal to 0.7, and the standard deviation of productivity shocks is equal to 0.1 as in Khan and Thomas (2013); Bloom, Guvenen, and Salgado (2023). Finally, we assume that innovations to firms' productivity at time t proportionally affect output at time t, so that $\sigma_y = \sigma_{\log z}$ and $\lambda_{yz} = 1$. In the baseline calibration, we set the exposure to idiosyncratic risk $\phi = 1$.

G.1 Numerical algorithm

Competitive Equilibrium. We use the separation result to compute the competitive equilibrium in two steps. First, we use the KFE and the good market-clearing condition to solve for the distribution of consumption shares $\eta_t(z)$ and for the wage w_t. Then, it is easy to compute aggregate labor and consumption using the labor market-clearing condition and the first-order condition for labor supply.

In order to solve for the distribution of consumption shares, we construct a grid for productivity $\mathbf{z}$ that takes values between $\underline{z}$ and $\bar{z}$. We denote by η the vector obtained by evaluating $\eta(z)$ at $\mathbf{z}$. Let $\mathbf{D_f}$ be the forward operator and $\mathbf{D_b}$ be the backward operator, defined as:

$$\mathbf{D_f} = \begin{pmatrix} -1/dz & 1/dz & 0 & \dots & 0 \\ 0 & -1/dz & 1/dz & 0 & \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \dots & & -1/dz & 1/dz \\ 0 & \dots & & 0 & -1 \end{pmatrix}$$

$$\mathbf{D_b} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1/dz & 1/dz & 0 & & \\ \vdots & -1/dz & 1/dz & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \dots & & -1/dz & 1/dz \end{pmatrix}$$

For first derivatives, we use the operator $\mathbf{D^1}$ defined as:

$$\mathbf{D^1} = \mathbf{1}(z > \bar{z})\mathbf{D_b} + \mathbf{1}(z < \bar{z})\mathbf{D_f}.$$

For second derivatives, we use the operator $\mathbf{D^2}$ defined as:

$$\mathbf{D^2} = \mathbf{D_f D_b}.$$

To handle the finite domain, we follow Moll (2014) to ensure that the support of the distribution is bounded. We assume two reflecting barriers at the bounds $\underline{z}$ and $\bar{z}$.

We initialize the algorithm with a value for the initial distribution of consumption shares $\eta_{\mathbf{0}}$. Then, at each step of the algorithm:

1. We use the good market clearing condition to find the wage that solves:

$$\left(\int_{w,}^{\infty} \frac{\eta_{zt}(z)}{\sigma_y(z)} z \frac{z - w_t}{\sigma_y(z)} dz \right) \frac{1}{\rho^e} = 1$$

2. We update the distribution $\eta_{\mathbf{1}}$ using the Runge-Kutta method

3. In order to minimize numerical errors when we iterate forward, we use our equation for $d\eta_e$ to normalize the distribution of consumption shares at each step, such that:

$${\eta_{\mathbf{1}}}^{norm} = \eta_{\mathbf{1}} \times \frac{\sum \eta_{\mathbf{0}} + d\eta_e}{\sum \eta_{\mathbf{1}}}$$

4. We update the distribution by setting $\eta_{\mathbf{0}} = {\eta_{\mathbf{1}}}^{norm}$ and iterate to convergence

We simulate the algorithm until the economy converges to a steady state, that is when $\eta_{\mathbf{1}} \approx \eta_{\mathbf{0}}$.

Once we have solved for the entire path of the distribution of consumption shares and wages, we can easily compute aggregate labor, consumption, and productivity as

$$Z_t = \left(\frac{1}{\rho^e} \int_{w_t}^{\bar{z}} \frac{z - w_t}{\sigma_y(z)} \frac{\eta_{zt}(z)}{\sigma_y(z)} dz \right)^{-1} \qquad \text{(Aggregate TFP)}$$

$$\ell_t^{1/\psi} = \left(\eta_{wt} Z_t \ell_t \right)^{-1} w_t \qquad \text{(Aggregate labor)}$$

$$c_t = \ell_t Z_t \qquad \text{(Aggregate consumption)}$$

Planner's Problem. The numerical algorithm that we use to solve the planner's problem is very similar to the one for the competitive equilibrium.

Once we have solved the entire path of the distribution of consumption shares and wages, as in the competitive equilibrium, we can easily find the optimal labor tax as a function of consumption shares, markups, and Pareto weights. Then, given the optimal labor tax and consumption shares, we solve for aggregate labor and consumption.

When we allow the planner to redistribute wealth at $t = 0$, we need to solve not only for the optimal labor tax at each t, but also for the distribution of consumption shares at $t = 0$. In order to solve this problem numerically, we parameterize the initial distribution of consumption shares. We follow Algan, Allais, and Den Haan (2008), and we approximate the distribution $\eta_0(z)$ with the exponential of a polynomial of degree N_p. This approximation turns an infinite-dimensional problem (i.e., the one of choosing an entire distribution) into a finite-dimensional problem, where the planner chooses coefficients $\rho_0^{N_p}, \rho_1^{N_p}, \ldots, \rho_{N_p}^{N_p}$ to maximize the objective function.

$$\eta_0(z) = \rho_0^{N_p} \exp\left(P(z, \rho^{N_p})\right)$$

In order to assess welfare gains, we need to evaluate the objective function of the social planner. We use the equation for the transformed planner's problem:

$$\begin{aligned}
&\int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log \ell_t - \Gamma_{wt} \frac{\ell_t^{1+1/\psi}}{1+1/\psi} dt + \int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log Z_t dt \\
&\quad + \int_0^\infty \Gamma_{wt} \log \eta_{wt} + \Gamma_{et} \log \eta_{et} dt \\
&\quad + \int_0^\infty \Gamma_{et} \left(\log \eta_{wt} - \log \eta_{et} + \left(\int_0^1 \log \eta_{i0} di - \log \eta_{w0} \right) + \frac{1}{\rho_e} \left(\rho_w - \rho_e + \frac{1}{2} \int_0^1 \sigma_{cit}^2 di \right) \right) dt
\end{aligned} \tag{G.2}$$

In principle, evaluating equation (G.2) poses some challenges, as it depends on the distribution of consumption shares across entrepreneurs at time $t = 0$ because of the term η_{i0}. In practice, we always consider situations in which the planner cannot alter the initial distribution of wealth between entrepreneurs with the same z in $t = 0$. Therefore, we rewrite equation (G.2) to obtain equation (G.3), where the term $\eta_{i0}/\eta_0(z_i)$ is the same in the competitive equilibrium and in the planner's allocation.

$$\begin{aligned}
&\int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log \ell_t - \Gamma_{wt} \frac{\ell_t^{1+1/\psi}}{1+1/\psi} dt + \int_0^\infty (\Gamma_{wt} + \Gamma_{et}) \log Z_t dt \\
&\quad + \int_0^\infty \Gamma_{wt} \log \eta_{wt} + \Gamma_{et} \log \eta_{et} dt \\
&\quad + \int_0^\infty \Gamma_{et} \left(\log \eta_{wt} - \log \eta_{et} + \left(\int_0^1 \log \frac{\eta_{i0}}{\eta_0(z_i)} di + \int \eta_0(z) f(z) dz - \log \eta_{w0} \right) + \right. \\
&\qquad \left. \frac{1}{\rho_e} \left(\rho_w - \rho_e + \frac{1}{2} \int_0^1 \sigma_{cit}^2 di \right) \right) dt
\end{aligned} \tag{G.3}$$

Finally, once we evaluate changes in welfare $\Delta Welfare$ between the planner allocation and the competitive equilibrium, we can easily evaluate welfare gains in consumption equivalent term by solving the following equation for x:

$$\Delta Welfare = \int_0^\infty \Gamma_{wt} \log(1 + x) + \int_0^\infty \Gamma_{et} \log(1 + x).$$