Wealth, Returns, and Taxation: A Tale of Two Dependencies*

Alexandre Gaillard†
Toulouse School of Economics

Philipp Wangner‡
Toulouse School of Economics

November 21, 2021
Job Market Paper [Latest version]

Abstract

We study wealth redistribution in a framework where individual portfolio choices and associated returns are correlated with wealth through: (i) *type* dependence, which reflects that investment skills drive return differences, and (ii) *scale* dependence, which captures that wealth itself triggers returns. Using an analytical framework, we argue that several common heterogeneous agent models can be understood through the lens of a type and scale dependence representation. We show that four key statistics characterize the macroeconomic and welfare implications of wealth taxation: the right tail of the wealth distribution, the degree of scale and type dependence, and the extent to which returns reflect investment productivity. We then build a quantitative model calibrated using micro US data and find an optimal marginal wealth tax rate of 0.8 percent above an exemption level of $550K. The result is driven by two opposing forces. Under scale dependence, productivity and wealth accumulation decrease with the tax, as risk-taking depends on wealth. Under type dependence, a higher wealth tax reinforces the selection of skilled investors at the top and improves productivity. Finally, the marginal wealth tax only slightly increases when returns partially reflect rent motives, as both forces almost quantitatively offset each other.

Keywords: Wealth taxation, Return heterogeneity, Type and scale dependence, Inequality.

*A previous version was entitled "*The Growth Irrelevance Frontier of Wealth Inequality*". We are extremely grateful to Christian Hellwig, Sumudu Kankanamge, Patrick Fève, Nicolas Werquin, Fabrice Collard and Tim Lee for helpful discussions and invaluable guidance. We also thank Andrew Atkeson, Adrien Auclert, Marianne Andries, George-Marios Angeletos, Tiziana Assenza, Saki Bigio, Timo Boppart, Pierre Boyer, Christophe Bruneel-Zupanc, Harold Cole, Mariacristina De Nardi, François de Soyres, Martial Dupaigne, Xavier Gabaix, Cécile Gaubert, Christian Gollier, Eugenia Gonzalez-Aguado, Alexander Guembel, Ulrich Hege, Nir Jaimovich, Thierry Magnac, Andreas Schaab, Ludwig Straub, Stéphane Villeneuve, Miguel Zerecero and participants at the 2021 SED, the 2021 EEA and various seminars for their constructive comments and discussions. The authors gratefully acknowledge funding from Banque de France. All remaining errors are our own.

†Email: alexandre.gaillard@tse-fr.eu. PhD candidate at the Toulouse School of Economics.

‡Email: philipp.wangner@tse-fr.eu. PhD candidate at the Toulouse School of Economics.

1 Introduction

Wealth is highly concentrated at the top. In the US, for instance, Saez and Zucman (2016) report that the wealth share of the richest 1% households has risen from 25% of the total wealth in 1980 to 40% in 2012. This trend has recently renewed academic and political interests on whether, and how, economies should redistribute the wealth of their richest individuals (see e.g. Piketty et al. (2014), Saez and Zucman (2019)). On the practical side, many of the richest OECD countries have considered a wealth tax in the last decades; some have implemented it, while others have withdrawn it.[1] In 2021, the Ultra-Millionaire Tax Act was proposed at the US Congress to introduce a tax on the wealth of the top 0.05% households.

In this paper, we investigate the macroeconomic and welfare implications of wealth taxation. To assess these implications, we build a general equilibrium model that accounts for key determinants behind the wealth accumulation of the richest individuals. We follow the influential work by Benhabib et al. (2011, 2019) and introduce heterogeneity in returns to wealth which constitutes, to date, one of the most compelling factors in explaining the high wealth concentration (Smith et al. (2019a), Hubmer et al. (2020), Xavier (2020)).[2] To do this, our model features two important departures from existing frameworks. First, in the spirit of Gabaix et al. (2016) we explicitly introduce two channels through which individuals differ in their returns to wealth: *type* and *scale* dependence. Type dependence reflects the fact that wealthy individuals obtain high returns because they differ in their innate or persistent characteristics, e.g. outstanding investment skills or high risk tolerance. Instead, scale dependence captures the fact that wealthier agents generate higher returns, regardless of their specific type, e.g. due to costly access to high-yield investments or decreasing relative risk aversion. This is especially relevant since recent contributions by Bach et al. (2020) and Fagereng et al. (2020) show that both concepts explain a substantial part of the cross-sectional correlation between returns to wealth and wealth. Second, we allow for the idea that private returns to wealth may only partially reflect differences in investment productivity due to some forms of rent-extraction. This may arise due to bargaining and market power, elite connections, or an unequal opportunity to access certain investments or markets (Piketty et al., 2014; Rothschild and Scheuer, 2016; Lockwood et al., 2017; Smith et al., 2019a).

We find that the optimal wealth tax rate in our benchmark model calibrated to the US economy is positive and large, at a rate of 0.8 percent above an exemption level of $550K. Our first key result is to show that this tax rate can be traced back to the underlying forces behind return heterogeneity and thus behind wealth accumulation. Specifically, the degree of type and scale dependence

[1] Abstracting from estate taxation and property taxes, twelve European countries levied an annual tax on net wealth in 1990. By 2018, only France, Norway, Spain, and Switzerland still imposed such a tax (Scheuer and Slemrod, 2021).

[2] Heterogeneity in the portfolio allocation of households is a commonly used factor to explain heterogeneity in returns to wealth (Calvet et al. (2019), Smith et al. (2019b), Meeuwis (2019), Xavier (2020)). Another factors that explain wealth concentration include, for instance, differential saving rates (Straub, 2018; Hubmer et al., 2020).

determines the sign and magnitude of the wealth tax. Under type dependence, a non-trivial selection effect of high-skilled investors at the top of the wealth distribution rationalizes a high wealth tax rate. In such a world, a wealth tax has the potential to increase overall productivity. In contrast, under scale dependence, a wealth tax substantially decreases wealth accumulation and productivity and provides a rationale for low wealth taxes or even a subsidy. Our second result shows that the optimal tax rate is surprisingly almost unresponsive to the extent to which returns to wealth reflect differences in investment productivity. When the strength of rent-extraction increases, the opposing forces arising from type and scale dependence almost exactly offset each other in our preferred benchmark calibration. We substantiate our quantitative results in two steps.

In a first step, we lay out the main concepts behind our results within an analytical two-period model. In this model, households' risk aversion is correlated with their initial wealth and their innate type, which determines their willingness to invest in risky but more productive assets. With a perfectly elastic supply of capital, aggregate productivity and output are determined in equilibrium by aggregating risky and riskless capital investments. We isolate and clarify the key parameters that characterize the macroeconomic and welfare implications of a change in top wealth inequality, due to, for instance, a wealth tax. We show that these implications depend on four statistics; (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking with respect to wealth, i.e. scale dependence, (iii) the sorting of individuals with different investment skill-types along the wealth distribution captured by the correlation between investor's types and wealth, i.e. type-dependence, and (iv) the extent to which returns to wealth reflect differences in productivity of investments rather than rents. While (i) – (iii) can in principle be measured empirically (Vermeulen, 2016; Bach et al., 2020; Fagereng et al., 2020), there is only little recent evidence concerning (iv) (cf. Lockwood et al. (2017) and Smith et al. (2019a)).

Despite its simplicity, the analytical model provides key insights regarding the role of inequality on aggregate output. We derive an intuitive diagram which captures all the possible relationships between changes in inequality and aggregate output or welfare resulting from the signs and the magnitude of scale and type dependence. We view this representation as a compelling device that can unify the existing literature studying, and disagreeing about, the relationship between inequality and output growth: a specific model can be classified in a particular region of our diagram given the underlying – implicit or explicit – assumptions regarding the above key parameters. For instance, models that incorporate mechanisms related to saving and investment decisions with a type and/or scale dependence representation (see among others Galor and Zeira (1993), Angeletos (2007), Cagetti and De Nardi (2006), Moll (2014), Gomez et al. (2016), Kaplan et al. (2018), Guvenen et al. (2019), Hubmer et al. (2020)) fall in a particular decomposition of our setup. This is especially important as scale dependence implies a strong behavioral response to a change in household wealth and thus makes aggregate responses considerably more sensitive to wealth inequality changes.

In our framework, the welfare-maximizing top wealth tax, based on a utilitarian consumption-equivalent variation welfare criterion, balances three effects. First, a marginal decrease in top wealth inequality through wealth redistribution affects productivity, output, and equilibrium wage rate, as it reallocates wealth among households who differ in their intrinsic investment skill or risk tolerance type and wealth. Second, whenever returns to wealth imperfectly reflect the productivity of investments, a change in inequality generates a change in the size of rents in the economy. That is, some individuals benefit from extra-returns, without actually affecting production. Therefore, aggregate returns adjust in equilibrium to ensure that the total capital income received by households equalizes the total product of capital redistributed in the economy. Third, redistribution from the top to the bottom induces a standard equity motive as wealth-rich and wealth-poor households differ in their marginal utility of consumption.

In a second step, we extend the framework to a full-blown quantitative model to carefully evaluate the implications of a wealth tax. The model is a variant of the standard incomplete-markets model with heterogeneous agents facing uninsurable labor income risk pioneered by Bewley (1986) – Huggett (1993) – Aiyagari (1994). Like in our simple model, heterogeneity in investment decisions and associated returns to wealth is introduced through type and scale dependence. In the spirit of Cagetti and De Nardi (2006), Moll (2014), and Benhabib et al. (2019), type dependence arises as households differ in their intrinsic ability to undertake risky productive investments. Importantly, this ability evolves stochastically but is highly persistent. The higher the persistence, the more likely high-skilled investors generate high returns during many periods, and the more frequently they are represented at the top of the wealth distribution. Furthermore, we incorporate two empirically relevant forms of scale dependence. First, conditional on being investors, richer agents invest a larger fraction of their wealth (intensive margin). Second, following Hurst and Lusardi (2004) or Fagereng et al. (2017), richer agents are more likely to become investors (extensive margin). Finally, on top of the possibility that rent-seeking may explain heterogeneity in returns to wealth (Rothschild and Scheuer, 2016), we also introduce elements that have been previously identified in the literature as having potential large implications for optimal capital taxation, i.e. we add a life-cycle structure with endogenous labor supply (Conesa et al., 2009; Kindermann and Krueger, 2014).

The model is calibrated to replicate the empirical labor income and wealth distributions and moments regarding the observed heterogeneity in portfolio choices across households. We separate risky but potentially more productive assets, such as private equity and public equity, from safe assets using estimates of returns from the PSID. Like Cagetti and De Nardi (2006) and Guvenen et al. (2019), heterogeneity in returns to wealth reflects investment productivity differences in our benchmark. We distinguish two types: highly skilled investors who manage a significant amount of risky equity assets, and non-investors who do not invest and constitute the vast majority of households in the SCF. As types are persistent, this approach generates an endogenous type dependence within the model. Second, following Hurst and Lusardi (2004), we exploit the panel dimension of the PSID to pin down scale dependence in the risky investment participation with respect to wealth such that it aligns with its empirical counterpart. Third, the share of wealth invested in risky equity, conditional on being an investor, is increasing along the wealth distribution. To distinguish scale dependence in the share arising from net risky investments only, we use detailed information from the SCF on the timing and the allocation of private equity business investments. Specifically, a large proportion of the increase in equity investments at the top is driven by recent additional private equity investments. To remain conservative, we only attribute this margin to scale dependence in the risky share invested, with the underlying assumption being that those additional investments are unlikely to drive the fortune of already rich households.

The benchmark model replicates the high concentration of returns at the top from both the type and scale dependence channels. To further investigate the properties of the model, we study alternative specifications with type or scale dependence only. We find that these alternatives are almost observationally equivalent to the benchmark model regarding the distributions of returns and wealth, i.e. both are able to generate high wealth concentration from persistent return heterogeneity. However, the aggregate responses to a wealth tax differ: the response is substantially amplified under a high degree of scale dependence.

Within a restricted class of wealth tax functions, we use our benchmark model to compute the long-run optimal one-time wealth tax reform, which we redistribute by lowering labor income taxes to obtain revenue neutrality. We jointly determine the marginal wealth tax rate and the exemption level above which it applies, which induces a common form of tax progressivity. Our result of an optimal tax of 0.8 percent above an exemption level of 550K is the first quantitative outcome of this setup. This reform generates a welfare gain equivalent to 0.14% of yearly consumption with large heterogeneity: they are high below the 70^{th} wealth percentile and negative at the very top. We then extend the model to account for the presence of rents in returns. Following evidence in Lockwood et al. (2017) and Rothschild and Scheuer (2016), we attribute the excess returns to wealth extracted from law and finance sectors to rent-seeking motives. Under this calibration and fixing the exemption at its benchmark level, the optimal marginal tax rate only slightly increases with the size of rents, to a rate of 0.92%. Dissecting our results, we find that they depend critically on whether top wealth inequality is driven by type or scale dependence.

If we first suppose the absence of rent extraction motives and scale-dependence is the dominant source of wealth concentration, then a higher wealth tax, by discouraging capital accumulation, causes a *snowball effect*: as agents become less wealthy, their rate of return falls, which further discourages productive investments. These self-enforcing effects imply that a wealth tax generates a large adverse behavioral response which decreases aggregate productivity. Instead, if we suppose that type-dependence is the dominant force at play, the wealth tax has a disproportionately large adverse effect on the investment of agents with high wealth but low returns, i.e. they dissave at a higher rate. Thus, the wealth tax creates an environment where *only the fittest survives at the top*, i.e. a selection effect whereby the top of the wealth distribution ends up being composed of the most productive investors. In this last case, although households accumulate less, the wealth tax has the property to raise aggregate productivity. In such a scenario, a wealth tax becomes a powerful instrument and, as shown by Guvenen et al. (2019), even superior to a capital income tax. In more concrete terms, fixing the exemption level at $550K, it is optimal to subsidy wealth with a negative tax rate of -0.8 percent in a model featuring scale dependence only. In contrast, it is optimal to heavily tax wealth at a rate of 2.4 percent in a model featuring type dependence only. In a world where both dependencies coexist, such as in our benchmark economy, the optimal tax rate falls in between those two bounds.

Now consider the case where high returns on wealth reflect rent extraction instead of more productive investment opportunities. In that case, the above conclusions are reversed. Under scale dependence, a wealth tax is desirable because it discourages inefficient rent-seeking behavior. Under type dependence, by contrast, the endogenous selection mechanism from the implementation of the wealth tax previously described still implies that agents with higher returns will be more concentrated at the top; but those higher returns are now a reflection of higher rents rather than higher productivity, thus making the wealth tax relatively undesirable. These two opposing forces rationalize a low response of the wealth tax rate to the size of the rent in the benchmark economy, as the two forces almost quantitatively offset each other.

Related literature Our work is related to a number of papers studying the relation between the distribution of wealth and its strong interplay with macroeconomic aggregates. Many macroeconomic models incorporate mechanisms related to saving and investment decisions with a type and/or scale dependence representation to generate realistic wealth distributions. For instance, Benhabib et al. (2019) construct a quantitative model designed to identify the determinants of wealth inequality and wealth mobility in the US. Their baseline model features wealth return heterogeneity due to type dependence only. Relatedly, Hubmer, Krusell and Smith Jr (2020) study how several determinants, comprising heterogeneity in wealth returns, account for the recent rise of wealth inequality in the US. They use estimates of returns from Bach et al. (2020) and interpret the observed heterogeneity in wealth portfolio and returns as scale dependence only. Other relevant examples include, among others, Cagetti and De Nardi (2006), Moll (2014), Kaplan et al. (2018), and Guvenen et al. (2019).[3] Yet, surprisingly, their systematic distinction has

[3] Among many others, Kaldor (1956), Stiglitz (1969), or Bourguignon (1981) study the role of wealth inequality in a neoclassical economy with convex scale dependence in saving behaviors. Other mechanisms include risk-taking

been neglected thus far. This paper fills this gap. We show that while both mechanisms are independently capable of generating large wealth inequality through return heterogeneity, they imply distinct macroeconomic and welfare implications from wealth redistribution.

Generally, our paper is related to a large literature quantifying optimal taxation in general equilibrium models with heterogeneity in household capital investments (see among others, Aiyagari (1995); İmrohoroğlu (1998); Kitao (2008); Conesa et al. (2009); Kindermann and Krueger (2014); Brüggemann (2020); Moll and Itskhoki (2019); Boar and Midrigan (2020)). Considering a tax on the stock of wealth, Shourideh et al. (2012) shows that a positive progressive tax on savings is optimal. Cagetti and De Nardi (2009) and De Nardi and Yang (2016) show that it is not welfare improving to abolish the estate tax in the US. Such taxes can be reinterpreted as intergenerational wealth taxes. Closer to our work is the recent contribution by Guvenen et al. (2019), which shows that heterogeneity in returns has the property to break the equivalence result between taxing capital flow relative to taxing the stock of capital under homogeneous returns. They find that replacing the capital income tax with a wealth tax reduces misallocation and increases overall welfare in a model in which heterogeneity in returns comes mainly from differential entrepreneurial skill-types. Our results point to the key role of type and scale dependence, together with whether returns reflect the productivity of capital, in deriving the welfare implications of a wealth tax. Empirically, Jakobsen et al. (2020) find a strong role of top wealth taxation on wealth accumulation. To our knowledge, two papers discuss the role of type and scale dependence for capital taxation in the spirit of Diamond (1998) and Saez (2001). Gerritsen et al. (2020) find that capital income tax is positive with returns heterogeneity under both dependencies. In complementary work, Schulz (2021) shows that the degree of scale dependence in returns significantly affects the capital income tax. Relative to them, we rather follow a different approach by studying optimal wealth taxation in a general equilibrium incomplete markets economy, and quantitatively show that it is critical to model the endogenous selection of investment skill-types along the wealth distribution.

Layout In section 2, we construct an analytical two-period version of our model to lay out the main concepts and forces at play. Section 3 sets out the quantitative dynamic model. Section 4 discusses the model's calibration and section 5 investigate its properties. In section 6, we use our model to study the welfare-maximizing wealth tax, and section 7 concludes the paper. The appendix contains all proofs, empirical analyses, and computational details.

behavior (Peress (2004), Brunnermeier and Nagel (2008), Calvet et al. (2009), Robinson (2012), and Meeuwis (2019)) non-convex investment cost and DRS (Banerjee and Newman (1993), Galor and Zeira (1993)), economies of scale in wealth management (Kacperczyk et al., 2019), social status derived from wealth holdings (Roussanov, 2010), investment in financial sophistication (Lusardi et al., 2017), wealth-dependent offshore investments (Alstadsæter et al., 2018), or non-convex investment costs in high return assets (Kaplan et al., 2018). In contrast, Kihlstrom and Laffont (1979), Moll (2014), Herranz et al. (2015), or Moll and Itskhoki (2019) introduce type dependence in which the distribution of types, and their persistence, is crucial to deriving aggregate efficiency. Combined dependencies arise in Quadrini (2000) and Cagetti and De Nardi (2006) through type dependence because entrepreneurs self-select at the top of the wealth distribution and through scale dependence due to wealth-driven occupational choices and a DRS technology.

2 An Analytical Two-Period Model

We begin with an analytical two-period framework to illustrate the conceptual distinction between type and scale dependence and the main trade-offs. The purpose of this section is to provide simple insights, and to introduce the notations used throughout the dynamic quantitative model.

2.1 Environment

Households A unit mass $i \in [0,1]$ of heterogeneous households lives for two periods, $t \in \{1,2\}$, with initial wealth a_0^i and innate risk-taking type ϑ^i drawn from the joint distribution $\mathcal{G}_0(\vartheta, a_0)$, with marginal distributions $g_\vartheta(\vartheta)$ and $g_{a_0}(a_0)$ defined over the support $\Theta \subset \mathbb{R}_+$ and $\mathcal{A}_0 \subset \mathbb{R}_+$. Households have CARA preferences over consumption c^i, i.e. $(1/\alpha^i)\left(1 - e^{-\alpha^i c^i}\right)$, where the absolute risk aversion α^i correlates with their initial wealth and risk-taking type, such that

$$\alpha^i \equiv \overline{\vartheta} \cdot \left[\vartheta^i (a_0^i)^\gamma\right]^{-1} , \tag{1}$$

where $\overline{\vartheta} \equiv \mathbb{E}[\vartheta]$ scales the average economy-wide risk tolerance. The parameter $\gamma \geq 0$ governs the shape of the household's risk tolerance in initial wealth. This preference specification captures in a reduced form various mechanisms driving type and scale dependence in portfolio choices and capital returns mentioned in the related literature.[4] In period $t = 1$, households invest optimally a share ω_1^i of their beginning of period wealth $a_1^i = a_0^i - t_a(a_0^i)$ into a *risky* innovative asset with stochastic gross return R_r^i, and the complementary share $(1 - \omega_1^i)$ into a *risk-free* asset with certain gross return R_f. The function $t_a(\cdot)$ defines a wealth tax on initial wealth and T is a second period lump-sum transfer. Agents inelastically supply one labor unit and obtain a wage w. In $t = 2$, returns and wage realize and households consume c_2^i. The objective of household i is given by

$$\max_{\{\omega_1^i\}} \left(1/\alpha^i\right)\left(1 - \mathbb{E}_1\left[e^{-\alpha^i c_2^i}\right]\right) \qquad \text{s.t.} \qquad c_2^i \leq \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i \omega_1^i\right) a_1^i + w + T , \tag{2}$$

where $\underline{r}$ is an aggregate return component, which is determined in equilibrium and common to all households.

Production In period $t = 2$, a competitive final good producer uses aggregate labor n and a continuum $j \in [0,1]$ of intermediate good projects x_s^j from two technologies $s \in \{N, I\}$, an innovative technology I with risky returns and a safe non-innovative technology N. The aggregate produc-

[4]This special form of CARA preferences extends the utility functions used in Alpanda and Woglom (2007) or Makarov and Schornick (2010) by specifying the wealth normalization with a power function. This form is also ultimately linked to Guiso and Paiella (2000) and Gollier (2001), who specify the shape of risk tolerance in terms of consumption rather than wealth.

tion function is given by $Y = Xn^{\varphi}$, where $X = \left(\sum_s \int_j x_s^j \, dj\right)$ and $\varphi \in [0, 1)$.[5] Profit maximization follows

$$\max_{\{n, \{x_s^j\}_{j,s}\}} Xn^{\varphi} - wn - \sum_s \int_j p_s^j \, x_s^j \, dj \, , \tag{3}$$

where p_s^j denotes the price of an intermediate good j in sector s.

An intermediate good producer uses risk-free capital k_N^j and risky capital k_I^j to run a project j with linear technologies. *Innovative* projects produce $x_I^j = (\phi\mu + A(1-\mu))k_I^j$, where $\phi > A$ denotes the expected innovate asset net return and $\mu \in [0, 1]$. *Traditional* projects operate with technology $x_N^j = Ak_N^j$. The revenue generated by a traditional safe project is $p_N^j x_N^j$ and the revenue from an innovative project is $p_I^j x_I^j$.

Market clearing The first order condition with respect to aggregate labor yields $w = \varphi X n^{\varphi - 1}$, with $n = 1$. Substituting for labor demand, the objective of the profit maximization (3) can be rewritten as $(1-\varphi)Xn^{\varphi} - \sum_s \int_j p_s^j \, x_s^j \, dj$. As intermediate goods are perfect substitutes, their prices are identical, and each unit is sold at a price $p_s^j = (1-\varphi)$.

The intermediate project i uses capital invested by household i to a run a project, such that $k_N^i = a_1^i(1-\omega_1^i)$ and $k_I^i = a_1^i \omega_1^i$. It redistributes revenues to household i as follows. Revenues from riskless assets are redistributed such that their returns equal the marginal product of capital net of wage payments, i.e. $R_f = (1-\varphi)A$. In contrast, the returns to innovative investments are given by $R_r^i = (1-\varphi)\kappa^i$, and may deviate from the net marginal product for two reasons. First, there is an idiosyncratic luck component $\kappa^i \sim \mathcal{N}(\phi, \sigma_\kappa^2)$ that introduces return risk on the household side. Second, we assume that there exists a return wedge between the expected risky return to wealth, $(1-\varphi)\phi$, and the net marginal product of innovative capital $(1-\varphi)\left(\phi\mu + A(1-\mu)\right)$, henceforth MPK_r, on the production side.[6] When $\mu = 1$, expected returns to innovative capital investments equal the MPK_r. Whenever $\mu < 1$, expected risky returns are higher than the MPK_r. In the extreme case where $\mu = 0$, the risk premium $(1-\varphi)(\phi - A)$ observed on the household side does not arise from productivity differences across asset classes.

A rationale for $\mu < 1$ comes from the presence of rent-extraction motives due to some forms of bargaining, market power or political connections of investors, i.e. in the words of Rothschild

[5] In Appendix OA 1.5 we derive the case with aggregate decreasing returns to scale in X. In this case the portfolio choice is increasing in X, as a higher X tends to depress the dispersion of returns. Therefore, the risky capital supply in this alternative model is upward-sloping.

[6] In our specification, idiosyncratic risk materializes as return risk on the investor side rather than idiosyncratic production risk. In this respect, our framework deviates from the seminal incomplete market growth economies of Angeletos and Calvet (2006); Angeletos (2007). We impose this assumption out of tractability. If the idiosyncratic capital income risk is modeled as an idiosyncratic productivity shock, one needs to integrate over the joint distribution of wealth and types to obtain aggregate output and productivity, similar to Gabaix (2011). In Appendix OA 1.2, we show that the policy functions are isomorphic in both cases. Aggregation follows under the additional assumption that there is a sub-continuum of agents in each state (ϑ, a_0).

and Scheuer (2016), "*the pursuit of personal enrichment by extracting a slice of the existing economic pie rather than by increasing the size of that pie*".[7] We view this return wedge as a stylized way to reconcile two approaches by acknowledging that empirically measured returns to wealth can not easily be partitioned into a rent component and the marginal product of capital. On the one hand, some work disentangles returns to wealth from MPK, either because of their partial equilibrium structure (Benhabib et al., 2019) or because of implicit full rent extraction (Hubmer et al., 2020). On the other hand, models with capitalists often assume a perfect pass-through between MPK and returns (see among others Cagetti and De Nardi (2006, 2009) or Guvenen et al. (2019)). Instead, we derive results for a range of values for the return wedge μ.

Under this structure, whenever $\mu < 1$, the aggregate return component $\underline{r}$ adjusts in equilibrium to ensure that the total product of capital generated on the production side coincides with the total capital income redistributed to the households by the intermediate good producer, such that

$$\underbrace{\int_i \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i\omega_1^i\right)a_1^i \, di}_{\text{returns}} = \underbrace{\int_i \left(A(1-\varphi)k_I^i + (\phi\mu + A(1-\mu))\,(1-\varphi)k_N^i\right) di}_{\text{capital product}} \,. \tag{4}$$

Distributional assumption Individual terminal wealth is affine in κ, i.e. its distribution is Gaussian $c_2^i \sim \mathcal{N}\left(\mu_{c_2}^i, \sigma_{c_2}^i\right)$ with mean $\mu_{c_2}^i = \varphi Y + T + (A(1-\varphi) + \underline{r})a_1^i + (1-\varphi)\omega_1^i a_1^i(\phi - A)$ and variance $\sigma_{c_2}^i = \left((1-\varphi)\omega_1^i a_1^i\right)^2 \sigma_\kappa^2$. Together with CARA preferences, this property ensures tractability of the equilibrium allocation. Finally, the initial wealth is assumed to be Pareto distributed.

Assumption 1 (Initial Wealth Distribution)**.** *Initial wealth is drawn from a Pareto law with scale $\underline{a}$ and shape $\eta > \max\{\gamma, 1\}$, such that $A_0 \sim \mathcal{P}a\left(\underline{a}, \eta\right)$ with $\mathbb{P}(A_0 \geq a_0) = (\underline{a}/a_0)^\eta$, $\forall a_0 \geq \underline{a}$.*

While theoretically convenient, this assumption is consistent with well-known empirical evidence documenting that the wealth distribution is right-skewed and displays an heavy upper tail (Vermeulen, 2016; Klass et al., 2006). The shape parameter η is inversely related to wealth inequality.[8] As changing η leads *ceteris paribus* to a change in aggregate wealth, we sometimes study the effect of varying η (*redistribution* effect) while preserving the same aggregate wealth level by adjusting the scale $\underline{a}$ (*level* effect).

2.2 Efficiency Gains and Redistribution

We begin by characterizing the inequality–efficiency trade-off in this economy. We focus on the aggregate allocation when investment decisions are driven by type and scale dependence and discuss wealth taxation at the end of this section. As such, we solve first for the case $t_a(a_0) = 0$.

[7] See notably the discussion in Scheuer and Slemrod (2020) and the work of Piketty et al. (2014) and Rothschild and Scheuer (2016). In a related paper, Boar and Midrigan (2019) study a setup with entrepreneurs and workers in which entrepreneurial returns to capital investment reflect partially market power.

[8] The Pareto tail is also inversely related to the wealth share $q(p)$ of the p wealthiest households by $q(p) = p^{1-1/\eta}$.

Lemma 1 (POLICY FUNCTIONS)**.** *Let us define* $\widetilde{\omega} \equiv \frac{\phi - A}{(1-\varphi)\sigma_\kappa^2}$ *such that the standard CARA risky asset share is* $(\widetilde{\omega}/\overline{\vartheta})$*. Under our extended preference form, household* i*'s risky asset share is given by*

$$\omega_1^i = \widetilde{\omega} \cdot \left(\vartheta^i/\overline{\vartheta}\right) \cdot (a_0^i)^{\gamma-1} . \tag{5}$$

If $\gamma = 1$ and $\vartheta^i = 1\,\forall i$, Lemma 1 provides the well-known result of Merton (1969) and Samuelson (1969), i.e. the share of risky asset holdings equals the baseline CARA solution captured by the risk premium over the variance of the risky return times the risk tolerance. Conditional on type ϑ^i, our CARA specification mimics IRRA (respectively DRRA) behavior if $\gamma < 1$ (respectively $\gamma > 1$). When $\gamma = 1$, our CARA specification nests CRRA behavior with constant risky asset share, while $\gamma = 0$ implies CARA behavior with constant risky asset holdings.[9] Therefore, the parameter γ pins down the elasticity of risky investments to initial wealth.

Using Lemma 1, we derive equilibrium quantities and prices.

Lemma 2 (AGGREGATE QUANTITIES)**.** *Given the joint distribution of types and wealth* $\mathcal{G}_0(\vartheta, a_0)$*, aggregate risky capital* K_I*, output* Y*, productivity* Z *and the wage rate* w *satisfy*[10]

$$K_I = \int_{\Theta\times\mathcal{A}_0} \omega_1(\vartheta, a_0) a_0 \, d\mathcal{G}_0(\vartheta, a_0) = (\widetilde{\omega}/\overline{\vartheta})\Big(Cov(\vartheta, a_0^\gamma) - \mathbb{E}\left[\vartheta\right]\mathbb{E}\left[a_0^\gamma\right]\Big) , \tag{6}$$

$$Y = Z\mathbb{E}\left[a_0\right] , \quad \textit{with} \quad Z = \mu(\phi - A)\frac{K_I}{\mathbb{E}\left[a_0\right]} + A , \tag{7}$$

$$\textit{and} \quad w = \varphi Y , \quad \underline{r} = (\mu - 1)(\phi - A)(1-\varphi)\frac{K_I}{\mathbb{E}[a_0]} . \tag{8}$$

Due to the CRS structure of final good production, demand for intermediate goods is perfectly elastic. Yet, its supply is bounded as households are risk-averse. Consequently, the risky portfolio shares of households, together with the joint distribution of wealth and types $\mathcal{G}_0(\vartheta, a_0)$, determine aggregate productivity Z and output Y. Therefore, wealth redistribution impacts productivity to the extent that it alters household i's investment in risky assets. The second condition of (8) states that $\mu < 1$ implies $\underline{r} < 0$, i.e. rent-extraction from risky investments induces a general equilibrium effect that decreases the common component of wealth returns, $\underline{r}a_1^i$, for all households.

We now formalize the effect of wealth redistribution on aggregate risky capital investment.

Proposition 1 (DISTRIBUTIONAL RELEVANCE)**.** *Consider without loss of generality a small mean pre-*

[9] Recall that one can always approximate the demand for risky assets for an arbitrary von Neumann-Morgenstern utility function as being proportional to risk tolerance, i.e. $\omega_1^i a_0^i \approx \frac{\mu_a^p}{var_a}\mathcal{T}(a_0^i)$. Contrary, under *generalized* CARA preferences risky investment is *exactly* proportional to our generalized risk tolerance $\mathcal{T}^i(\vartheta^i, a_0^i) = (\vartheta^i/\overline{\vartheta})(a_0^i)^\gamma$.

[10] The second condition in (8) follows by assuming that there is a sub-continuum of households in each state (ϑ, a_0^i).

serving change in the Pareto tail $\eta' > \eta$. Its effect on aggregate risky capital K_I can be decomposed into

$$\Delta K_I(\eta', \eta) = \underbrace{\Delta^a K_I(\eta', \eta)}_{\substack{\text{scale dependence in}\\ \text{portfolio holdings}}} + \underbrace{\Delta^\vartheta K_I(\eta', \eta)}_{\substack{\text{type heterogeneity and}\\ \text{selection}}} \quad ,$$

where $\Delta^a K_I(\eta', \eta)$ is zero if $\gamma \in \{0, 1\}$, increasing in η' if $\gamma \in (0, 1)$ and decreasing in η' if $\gamma > 1$. A sufficient condition for $\Delta^\vartheta K_I(\eta', \eta)$ to decrease in η' is $\left(\frac{\partial corr(\vartheta, a_0^\gamma)}{\partial \eta} + \frac{\partial corr(\vartheta, a_0^\gamma)}{\partial \underline{a}} \frac{\underline{a}}{\eta(\eta-1)} \right) \frac{1}{corr(\vartheta, a_0^\gamma)} \leq 0.$

Proposition 1 establishes general conditions under which the wealth distribution is a *relevant* equilibrium object by decomposing the effect of a change in the tail of the wealth distribution, η, on aggregate risky capital K_I into two terms: (i) a scale dependence term $\Delta^a K_I(\eta', \eta)$, which hinges on the risk taking elasticity γ, and a (ii) type dependence term $\Delta^\vartheta K_I(\eta', \eta)$, which encapsulates the selection of ϑ-types across the wealth distribution. A change in the Pareto tail is called *distributional relevant* if both effects do not offset each other.

In the absence of type dependence, e.g. $\vartheta^i = \overline{\vartheta} \, \forall i$, wealth redistribution from the top to the bottom decreases (respectively increases) K_I if $\gamma > 1$ (respectively $0 < \gamma < 1$). When $\gamma = \{0, 1\}$, distributional *irrelevance* arises as aggregate variables do not depend on the distribution of wealth, either because risky investments are constant ($\gamma = 0$), or because the share invested is constant ($\gamma = 1$). For the sake of clarity, we refer to scale dependence as a situation in which $\Delta^a K_I(\eta', \eta) \neq 0$, while a negative (respectively positive) scale dependence corresponds to the case where $\Delta^a K_I(\eta', \eta) > 0$ (respectively $\Delta^a K_I(\eta', \eta) < 0$). Our notion of scale-dependence therefore corresponds to cases in which scale effects, through γ, generate a *distributional relevant* link between wealth redistribution and aggregate quantities. This arises when the individual portfolio policy function ω_1^i depends non-linearly on initial wealth.

In the presence of type selection, the sufficient condition in Proposition 1 provides the bound on the change in the correlation between innate types and initial wealth such that $\Delta^\vartheta K_I(\eta', \eta) < 0$. Therefore, even if $\gamma = 1$, the distribution of wealth may be *relevant* through type dependence.

2.2.1 The Efficiency-Inequality Decomposition: A Closed-form Representation

Although Proposition 1 is general, we subsequently study a tractable representation of the equilibrium and the effects of wealth redistribution by putting a structural assumption on $Cov(\vartheta, a_0^\gamma)$.

Assumption 2 (Joint Distribution)**.** *Let $\vartheta \sim \mathcal{P}a(\underline{\vartheta}, \epsilon)$ such that $\overline{\vartheta} = \frac{\underline{\vartheta}\epsilon}{(\epsilon - 1)}$. The joint cdf $\mathcal{G}_0(\vartheta, a_0)$ is constructed based on the Farlie-Gumbel Morgenstern copula with dependence parameter $\varrho \in [-1, 1]$.*

Under Assumption 2, when $\varrho > 0$ (respectively $\varrho < 0$), there is a positive (respectively negative) correlation of types and wealth, while $\varrho = 0$ induces no correlation.[11] The level of ϱ translates

[11] The dependence parameter ϱ and the Spearman's correlation, ϱ^s, are under the Farlie-Gumbel Morgenstern copula related by $\varrho^s = \varrho/3$.

into the degree of selection, which is, for simplicity, exogenous in this section.

The following result decomposes the trade-off between inequality and efficiency into four terms which capture type dependence, scale dependence, an interaction term and the extent to which the excess return to wealth reflects productivity differences.

Proposition 2 (Efficiency-Inequality Relation)**.** *Given Assumptions 1-2 and an aggregate positive riskless capital supply in equilibrium, i.e.* $\frac{K_I}{\mathbb{E}[a_0]} < 1$*, wealth-normalized output is given by* $\widetilde{Y}(\eta) = A + \mu(\phi - A)\widetilde{\omega}\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\eta-1}{\eta-\gamma}\underline{a}^{\gamma-1}$.[12] *The marginal effect of wealth redistribution on* $\widetilde{Y}(\eta)$ *is*

$$\frac{\partial\widetilde{Y}(\eta;\gamma,\varrho)}{\partial\eta} \propto -\mu(\phi - A)\Big(\underbrace{\Omega^{\gamma}(\eta,\gamma)\cdot(\gamma-1)}_{\textit{scale dependence}} + \underbrace{\Omega^{\varrho}(\eta,\gamma)\cdot\varrho}_{\textit{type dependence}} + \underbrace{\Omega^{\varrho\gamma}(\eta,\gamma)\cdot\rho(\gamma-1)}_{\textit{interaction term}}\Big), \tag{9}$$

where $\Omega^{\gamma}(\eta,\gamma)$, $\Omega^{\varrho}(\eta,\gamma)$ *and* $\Omega^{\varrho\gamma}(\eta,\gamma)$ *are strictly positive inequality multipliers.*

A key property of Proposition 2 is the ambiguous effect of rising wealth inequality on normalized output that depends on the relative strength of scale dependence, type dependence and their interaction, respectively captured by the terms $\gamma - 1$, ϱ and $\varrho(\gamma - 1)$. If preferences mimic DRRA behavior ($\gamma > 1$) and there is a positive selection of types ($\varrho > 0$), then output unambiguously rises in response to higher wealth inequality, since it reallocates wealth to agents investing in riskier and more productive assets ($\mu > 0$). This effect is scaled to the degree to which returns to investment reflect differential capital productivity, captured by the term $\mu(\phi - A)$.

Importantly, variations of the Pareto tail exhibit highly nonlinear effects on output captured by the inequality multipliers $\Omega^{\gamma}(\eta,\gamma)$, $\Omega^{\varrho}(\eta,\gamma)$ and $\Omega^{\varrho\gamma}(\eta,\gamma)$. In practice, the precise decomposition and the associated inequality multipliers are model-specific; however, as discussed below, the general idea and mechanisms unify a number of frameworks. In complete unequal economies, i.e. $\eta \to \max\{\gamma, 1\}$, small variations of the Pareto tail result in rather large output variations. In contrast, in a complete egalitarian societies, i.e. $\eta \to \infty$, small variations in η result in small output variations. Intuitively, in more unequal economies, scale dependence and selection effects are stronger in magnitude such that even small variations of η lead to strong investment reallocations.

Equation (9) also implies that, for a given level of inequality η, there exists an infinite number of possible combinations of type and scale dependence on a bounded two-dimensional set consistent with a given marginal effect of wealth redistribution on output. In Definition 1, we specify the notion of *iso-growth* (a kind of isoquant) which describes all parameter pairs (γ, ϱ) for which a marginal variation of the Pareto shape η generates a given output response $\overline{g}$.

Definition 1 (Iso-growth of inequality)**.** *For a given wealth Pareto tail* η *and* $\mu > 0$*, the iso-growth at level* $\overline{g}$ *is defined by the pair* (γ, ϱ) *that satisfies* $isoG(\eta, \overline{g}) \equiv \left\{(\gamma,\varrho) \in \Gamma \times [-1, 1] : -\frac{\partial\widetilde{Y}(\eta;\gamma,\varrho)}{\partial\eta} = \overline{g}\right\}$.

[12]The condition $\frac{K_I}{\mathbb{E}[a_0]} < 1$ is satisfied for a given $\underline{a}$ if $\frac{\eta-\gamma}{\eta-1} \geq \widetilde{\omega}\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\underline{a}^{\gamma-1}$.

A special case ensues for $\overline{g} = 0$ for which the *iso-growth* curve separates the *growth enhancing region*, i.e. $-\frac{\partial \widetilde{Y}(\eta;\gamma,\varrho)}{\partial \eta} > 0$, from the *growth dampening region*, i.e. $-\frac{\partial \widetilde{Y}(\eta;\gamma,\varrho)}{\partial \eta} < 0$. For this reason, we label this special *iso-growth* curve the *Growth Irrelevance Frontier* of wealth inequality. Lemma 6 in Appendix A.1.7 provides conditions for its existence based on the strength of type and scale dependence and the wealth Pareto tail η. If these parameter restrictions do not apply, an infeasible pair $(\gamma, \varrho) \notin \Gamma \times [-1, 1]$ would be required to obtain *growth neutrality*.[13] On the *Growth Irrelevance Frontier (GIF)* of wealth inequality, type and scale dependence exactly offset each other or are absent. In Lemma 3, we provide key properties of the *iso-growth*; it is decreasing in the space (γ, ϱ) and rotates clockwise in the wealth Pareto shape η.

Lemma 3 (PROPERTIES OF THE GIF AND ISO-GROWTH)**.** *The GIF is strictly decreasing on the defined set of Lemma 6, i.e.* $\frac{\partial \gamma}{\partial \varrho}|_{d\eta=0} < 0$*. A higher tail* η *rotates the GIF such that* $\frac{d\gamma}{d\eta}|_{d\varrho=0} > 0$ *for* $\gamma > 1$ *and* $\frac{d\gamma}{d\eta}|_{d\varrho=0} \leq 0$ *for* $\gamma \leq 1$*. Also, for a higher level* $\overline{g}$*,* $isoG(\eta, \overline{g})$ *is the translation of the GIF and* $\frac{d\gamma}{d\overline{g}}|_{d\varrho=0} > 0$*.*

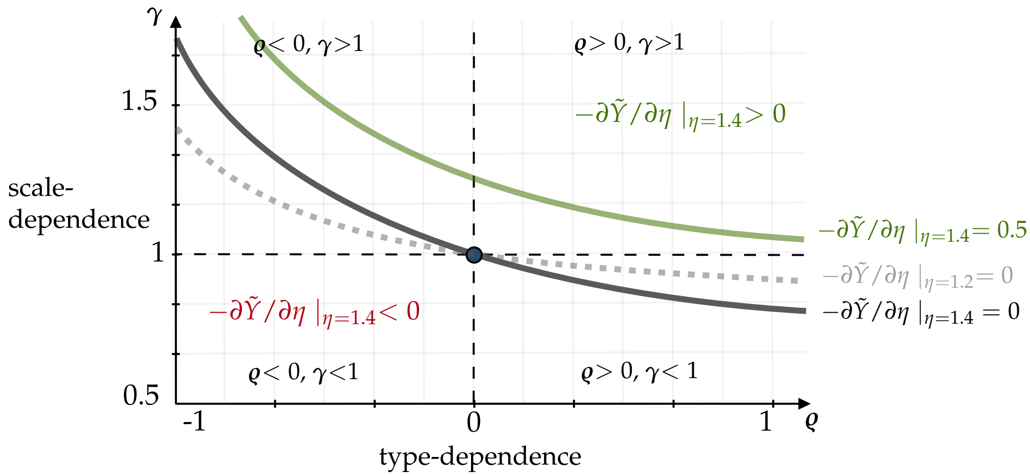
Figure 1 illustrates the *iso-growth* for two different Pareto tails η and output levels $\overline{g}$. There are four regions which are delimited by the sign of type and scale dependence. In the top-right and bottom-left regions, the wealth-dependent risk-taking and the selection effect move in the same direction. An increase in inequality therefore unambiguously induces more (respectively less) economy-wide risk-taking and higher (respectively lower) productivity. In the top-left and bottom-right regions, the wealth-dependent risk-taking and the selection effect move in opposite directions. In these regions, there exist infinite combinations of type and scale dependence such that both effects offset each other giving rise to the GIF. In the top-left region characterized by $\{\varrho < 0, \gamma > 1\}$, an increase in inequality leads to higher output only if the positive scale dependence is sufficiently strong. In contrast, in the bottom-right region characterized by $\{\varrho > 0, \gamma < 1\}$ an increase in inequality decreases output if the wealth-dependent risk-taking elasticity γ is sufficiently low for a given selection $\varrho > 0$.

For a higher effect of wealth inequality on the level of output, i.e. an increase in $\overline{g}$, the iso-growth moves upward in the (γ, ϱ) diagram; the higher effect of greater wealth inequality on output can only be rationalized with higher degrees of positive type and/or scale dependence. Finally, the level of wealth inequality changes the relative strength of type and scale dependence effects; a higher inequality (lower η) reinforces the scale dependence effect relative to the selection effect and, as a result, the GIF (and the translated *iso-growth* curve) flattens. Finally, for $\overline{g} > 0$ (respectively $\overline{g} < 0$), a given $isoG(\overline{g}, \eta)$ shifts downward (respectively upward) with an increase in the productivity gap $\mu(\phi - A)$ as less reallocation between the two productive sectors is needed to achieve a certain level $\overline{g}$.

[13] Notice that the $isoG(\eta, 0)$ does not exist in a complete unegalitarian economy ($\eta = \max\{1, \gamma\}$), since in this case the absolute strength of scale dependence dominates the selection effect. However, in relatively egalitarian economies, both effects are small in magnitude such that the $isoG(\eta, 0)$ exists on a bounded set.

Figure 1. The inequality–efficiency diagram.

Note: numerical parameter values are $\varepsilon = 2.0, \underline{a} = \underline{\vartheta} = 1.0, A_r = 1.1, \sigma_\kappa = 0.2, A = 1.0$.



Legend: The solid black line is the GIF with a Pareto tail $\eta = 1.4$, and the dotted grey line is the GIF with $\eta = 1.2$. The solid green line is an iso-growth curve corresponding to a growth level of 0.5 percent.

Despite its simplicity, the analytical decomposition derived above allows to captures key mechanisms driving real-world household behavior, and carries over different quantitative models with heterogeneity in household investments such as the one studied in section 3. It shows that, in order to understand the effects of wealth redistribution on aggregates in a framework that accounts for heterogeneous capital investments, there are four key parameters required: (i) the Pareto tail of the wealth distribution, η, (ii) the elasticity of risk-taking with respect to wealth, γ, and (iii) the selection or sorting of types along the wealth distribution, as captured by the dependence, ϱ, and (iv) the excess wealth return augmented by the extent to which higher returns reflect higher capital investment productivity, $\mu(\phi - A)$.

Discussion and practical implications A number of frameworks with type and scale dependent mechanisms can be unified within the representation of the inequality-efficiency diagram of Figure 1 and equation (9).[14] In the basic Aiyagari (1994) economy, the distribution of wealth is almost growth–irrelevant due to the quasi linearity of the decision to save of the wealth-rich households. Angeletos (2007) studies an economy with linear portfolio policy functions and no type heterogeneity, thereby implying distribution *irrelevance*. Those models are located respectively at, and close to, the anchor point of the inequality-efficiency diagram ($\gamma = 1$ and $\varrho = 0$). Models with capitalists/entrepreneurs (among others Cagetti and De Nardi (2006, 2009), Guvenen et al. (2019), Brüggemann (2021)) often display positive type dependence at the stationary equilibrium, as entrepreneurs who invest in high-return private equity investments self-select at the top of the wealth distribution ($\varrho > 0$). However, they feature decreasing marginal product

[14] In the Online Appendix OA 1.1, we hypothetically locate various theoretical and quantitative incomplete markets models relative to the GIF.

on those investments which can be reinterpreted as negative scale dependence effects ($\gamma < 1$). In those models, wealth redistribution from the top to the bottom has thus an ambiguous effect on output. Therefore, they are positioned in the bottom-right region. Similarly, the seminal paper by Galor and Zeira (1993) with non-convex human capital investment costs can be reinterpreted as a form of DRS ($\gamma < 1$) but without type heterogeneity ($\varrho = 0$). Finally, the model of Moll (2014) displays only type dependence in households capital productivity with linear investment policy functions. Such models therefore locate on the locus with $\gamma = 1$.[15]

From the above diagram, it is interesting to see that type and scale dependence cannot be simply identified by using information on the effect of inequality on growth, as an infinite combination of pairs (γ, ϱ) may rationalize the relationship.[16] To identify both dependencies, it is ideal to have access to detailed panel data comprising portfolio decisions and associated returns to wealth. This may for example allow an econometrician to infer individual types through fixed effects, and scale dependence by estimating the effects of wealth variations on household behavior. Recent papers, such as Fagereng et al. (2020) and Bach et al. (2020), pave the way for such an empirical analysis. Without access to panel data and by relying only on cross-sectional data it is however difficult to identify the distinction as both channels may in principle generate consistent patterns regarding the portfolio allocation and associated returns to investment along the wealth distribution.[17] This is striking given that both effects result in substantially different elasticities of macroeconomic aggregates to wealth redistribution and, as shown in the dynamic model of section 3, to distinct implications for optimal wealth taxation.

A numerical example Consider two distinct models, i.e. the first with scale dependence only, and the second with type dependence only. We normalize the non-innovative productivity $A = 1$ and assume $\mu = 1$. We set the labor share $\varphi = 0.67$, the shape $\eta = 1.4$ consistent with estimates for the US (Vermeulen, 2016), and the risk premium such that $(1 - \varphi)(\phi - A) = 15\%$. The variance σ_κ^2 is set to 0.16, which is consistent with estimates from the PSID discussed in section 3. We calibrate type and scale dependence to generate a consistent cross-sectional pattern of risky equity shares,

[15]In Appendix OA 3.1, we show how a number of additional frameworks can be understood within this representation. Moreover, the representation above provides a rationale for the lack of clear empirical evidence on the role of wealth inequality on growth (see for instance Perotti (1996); Forbes (2000); Barro (2008)). Recent panel data estimations tend to find a weak *positive* relationship in developed countries (Forbes, 2000; Voitchovsky, 2005; Barro, 2008; Frank, 2009) and a *negative* one in developing economies. The disparity of the estimates can be reconciled within our framework as the relation crucially depends on the underlying selection of agents, as well as the direction of scale dependence, which may of course be country-specific. In the Online Appendix OA 3.1, we investigate the relationship between top wealth inequality and GDP growth using new estimates regarding wealth concentration and find a positive link.

[16]They are, however, identified under particular conditions. Figure 1 shows that γ and ϱ are identified with *two* different couples of observation $(\eta_1, \overline{g}_1)$ and $(\eta_2, \overline{g}_2)$, but this requires that type and wealth dependence are constant over time. Moreover, estimating this relationship is somewhat complex as shown by the variety of empirical results in this related literature. See among others Forbes (2000), Voitchovsky (2005) or Barro (2008).

[17]See Section 3 for a detailed discussion regarding the identification of type and scale dependence in micro datasets.

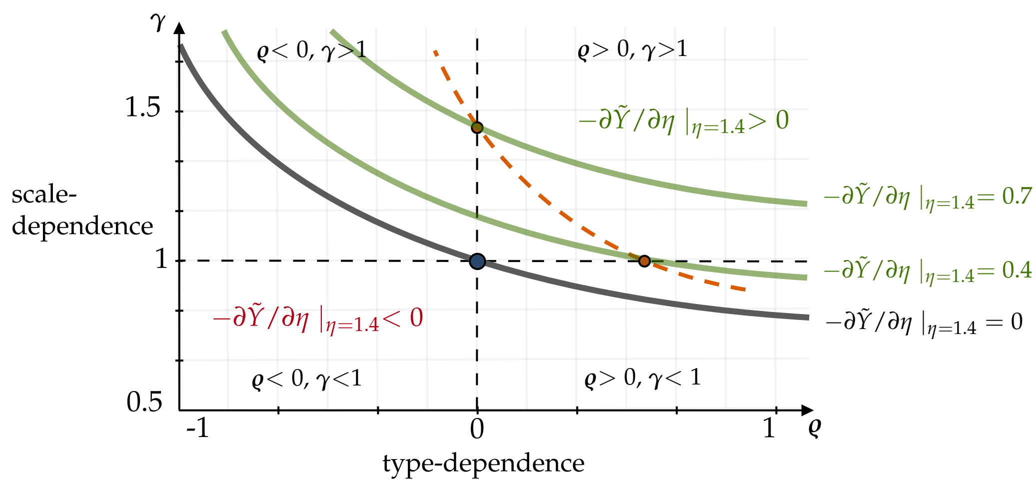
thus targeting the equity share of the top 1% wealthiest households of 65% as observed in the 2010 SCF. Note that because the average risky equity share is increasing in wealth in the cross-section, the two models are located respectively on the $\{\varrho = 0, \gamma > 1\}$ and $\{\varrho > 0, \gamma = 1\}$ *loci*.

In the type dependence model, we set the Pareto shape of types $\varepsilon = 2$ and vary the correlation between types and wealth to match the top 1% risky share, such that $corr(\vartheta, a_0) = 0.65$.[18] In the scale dependence model, we match the same target by varying the wealth-dependent risk-taking elasticity and obtain $\gamma = 1.39$. In Appendix A.2, we show that both models reproduce well the overall cross-sectional pattern of portfolio shares, at the bottom and at the upper end of the wealth distribution. However, the responses of output to a proportional top marginal wealth tax of 1% on the top 1% wealthiest households, which is redistributed through lump-sum transfers, differ substantially. We find that output drops by 0.43% under type dependence, which is in effect substantially lower than the 0.70% reduction found under scale dependence. In Figure 2, we report their respective iso-growth location. The difference in output responses originates from the behavioral response triggered by scale dependence as wealth varies (cf. Lemma 1).

This simple numerical example illustrates the importance of unraveling the economic forces behind capital investment heterogeneity. Notice that in a dynamic model, scale effects will be further amplified, as future wealth is a function of current investment itself, and the selection of skill-types along the distribution will be endogenous to inequality changes.

Figure 2. The inequality–efficiency diagram.

Note: numerical parameter values are $\varepsilon = 2.0, \underline{a} = \underline{\vartheta} = 1.0, A_r = 1.1, \sigma_\kappa = 0.2, A = 1.0$.



Legend: the solid black line is the GIF with an inequality level $\eta = 1.4$. The solid green lines are iso-growth curves corresponding to the two orange dots illustrating the numerical example described in the main text. The orange dashed line represents all combinations (ϱ,γ) such that the model replicates the cross-sectional portfolio shares and returns in the data, for a given $\eta = 1.4$.

[18]Instead, it is also possible to fix the correlation between types and wealth but vary the extent to which individuals are different by varying the shape ϵ. There is marginal difference in considering one over the other alternative, as long as the cross-sectional distribution of portfolio shares and wealth are well matched.

2.3 From Efficiency to Welfare

We now shift our focus to a welfare analysis. For tractability, we consider the case of a constant rate of progressivity (CRP) tax (Feldstein, 1969; Heathcote et al., 2017) on initial wealth such that $t_a(a_0^i) = a_0^i - (a_0^i)^{1-p_a}$, where $p_a \in (-\infty, 1)$ captures the progressivity of the wealth tax schedule. We denote with "$\sim$" *post-tax* variables. Under this assumption, individual choices are isomorphic, replacing initial wealth a_0^i with $\widetilde{a}_1^i = a_0^i - t_a(a_0^i)$, which implies an updated first period wealth Pareto tail $\widetilde{\eta} = \frac{\eta}{1-p_a}$ and scale $\widetilde{\underline{a}} = \underline{a}^{1-p_a}$, where $p_a \to 1$ implies a complete egalitarian economy. We measure welfare in terms of consumption equivalents, defined as the amount $\Delta^{CE,i}$ that makes household i in the reformed economy as well off as in the initial status quo economy, such that $\mathbb{E}[u(\widetilde{c}_2^i - \Delta^{CE,i})] = \mathbb{E}[u(c_2^i)]$. Under our CARA-Normal structure, this gives $\Delta^{CE,i} = \widetilde{x}_{c_2}^i - x_{c_2}^i + \frac{\Delta_c^i}{\widetilde{\alpha}_i}$, where $x_{c_2}^i$ and $\widetilde{x}_{c_2}^i$ denote certainty equivalents of the second period pre- and post-tax consumption, and the term Δ_c^i arises as the utility is a positive function of initial wealth through the risk aversion α^i. Given an utilitarian equivalent variation-based welfare measure, the planner solves

$$\mathcal{W} = \arg\max_{p_a} \int \Delta^{CE}(\vartheta, a_0)\, d\mathcal{G}_0(\vartheta, a_0) \qquad \text{s.t.} \qquad T = \frac{\eta \underline{a}}{\eta - 1} - \frac{\eta \underline{a}^{1-p_a}}{\eta - 1 + p_a}\,. \tag{10}$$

where the last equality balances the government budget constraint, such that $\int t_a(a_i^0) di = T$.[19]

Lemma 4 (Optimal wealth redistribution)**.** *Assume that the excess return is sufficiently large, i.e.* $\phi - A - \frac{\sigma_\kappa^2}{2}(\widetilde{\omega}/\overline{\vartheta}) > 0$. *The optimal progressivity* p_a^* *solves* $\frac{\partial \mathcal{W}}{\partial p_a} = 0$, *with*

$$\frac{\partial \mathcal{W}}{\partial p_a} = \Bigg(\underbrace{\varphi \frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}} \mathbb{E}[a_0]}_{\substack{\text{GE wage} \\ \text{efficiency if } \mu > 0}} + \underbrace{\frac{\partial \underline{r}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}} \mathbb{E}[\widetilde{a}_0]}_{\substack{\text{GE rent component} \\ \text{if } \mu < 1}} \Bigg) \frac{\partial \widetilde{\eta}}{\partial p_a} + \underbrace{\frac{\partial T}{\partial p_a}}_{\substack{\text{lump-sum} \\ \text{transfers}}} + \underbrace{\int \mathcal{R}(a_0, \vartheta) d\mathcal{G}_0(a_0, \vartheta)}_{\text{direct effects of the wealth tax}}$$

where $\mathcal{R}(\vartheta, a_0)$ *captures the direct effects of the wealth tax on second period consumptions and on risk aversion* $\widetilde{\alpha}_i$. *The sign of* $\frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}$ *is characterized in Proposition 2 and* $sgn\left(\frac{\partial \underline{r}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}\right) = -sgn\left(\frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}\right)$.

Lemma 4 characterizes the main trade-offs of wealth taxation on welfare. First, there is an equity channel as individuals differ in initial wealth, and thus in their marginal utility of consumption. Welfare increases with the lump-sum transfers T, but decreases with the efficiency losses from wealth taxation. The latter affects terminal wealth, captured by the term $\mathcal{R}$, through behavioral investment responses, i.e. changes in ω_1^i and changes in the curvature of the utility function through the risk aversion α^i. Second, a wealth tax affects efficiency, captured by $\frac{\partial \widetilde{Y}}{\partial \widetilde{\eta}}$, by reallocating wealth from the top to the bottom of the wealth distribution. Depending on the size of type and scale dependence, this affects the amount of capital that is invested in the innovative

[19] Between periods $t = 0$ and $t = 2$, the government may in principle invest tax revenues in risky or riskless assets, K_N or K_I, and obtain returns from those investments which affect the amount of second period lump-sum transfers T. We simplify the exposition here and assume that the government does not invest.

sector, and thus aggregate productivity and wage w. Third, whenever $\mu < 1$, the size of the rents in the economy responds to the inequality change induced by the wealth tax, captured by the aggregate return component $\underline{r}$. If rent-extraction increases, it lowers welfare as high capital return investors obtain a larger fraction of the overall product of capital, such that risky private returns to wealth become larger than their social value, i.e. the marginal product of capital investments. This in turns lowers the aggregate capital return component $\underline{r}$ to ensure that the total product of capital equalizes the amount of returns redistributed to households. In other words, the existence of rents widens the dispersion of returns to wealth across households, without actually reflecting dispersion in investment productivity.

The optimal proportional wealth tax thus trades-off equity and rent-extraction *versus* efficiency considerations. In this static model with exogenous type dependence, a lower μ implies a higher progressivity of the wealth tax. In the quantitative dynamic model in which the joint distribution of skill-types and wealth is endogenous, we argue that this result crucially depends on how the selection of skilled investors reacts to the implementation of a wealth tax in the long-run.

2.4 Generalization

The assumptions made throughout the special case allowed to isolate risk-taking decisions arising from *type* and *scale* dependence. We now briefly extend the analysis to saving decisions, alternative sources of *scale* dependence, and aggregate productivity shocks.[20]

2.4.1 Portfolio choice with explicit saving decision

Households now consume over the two periods, c_1^i and c_2^i, and preferences have a recursive form $u_1^i = U(c_1^i) + \beta U\Big(G^{-1}\Big(\mathbb{E}\Big[G\Big(U^{-1}(u_2^i)\Big)\Big]\Big)\Big)$, where it holds that $u_2^i = U(c_2^i)$, $U(c^i) = \frac{c^{1-1/\sigma}}{1-1/\sigma}$ and $G(c^i) = (1/\alpha^i)\left(1 - e^{-\alpha^i c^i}\right)$. As a result, the maximization objective becomes

$$\max_{\{c_1^i,\, \omega_1^i,\, a_1^i \geq 0\}} \frac{1}{1-1/\sigma}\left((c_1^i)^{1-1/\sigma} + \beta\left\{-\frac{1}{\alpha^i}\log\left(\mathbb{E}\left[e^{-\alpha^i c_2^i}\right]\right)\right\}^{1-1/\sigma}\right), \tag{11}$$

$$\text{s.t.} \qquad c_1^i + a_1^i \leq a_0^i - t_a(a_0^i)\,, \qquad c_2^i \leq \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i\omega_1^i\right)a_1^i + w + T\,,$$

where $\sigma > 0$ and $\beta \in (0,1)$ define, respectively, the intertemporal elasticity of substitution (IES) and the discount rate. In this case, the joint heterogeneity in marginal propensities to save and marginal propensities to take risk (Kekre and Lenel, 2020) is key to studying aggregate allocations and gives rise to generalized *iso-growth* curves. For simplicity, we neglect wealth taxation in the following, i.e. $t_a(a) = T = 0$.

[20] Additional extensions include uninsurable labor income risk and participation decisions. We relegate those additional analyses to the online appendix OA 1.8.

Lemma 5 (INDIVIDUAL PORTFOLIO CHOICE). *Denote $\widetilde{\beta} = (\widetilde{R}\beta)^{\sigma} + \widetilde{R}$ where $\widetilde{R} = \underline{r} + R_f$ and let us assume an interior solution to (11) such that $c_1^i > 0$. Individual portfolio choices of risky and riskless assets are denoted, respectively, $k_1^i = \omega_1^i a_1^i$ and $b_1^i = (1-\omega_1^i)a_1^i$, and given by*

$$k_1^i = \widetilde{\omega}\frac{\vartheta^i}{\overline{\vartheta}}(a_0^i)^{\gamma}\ , \quad b_1^i = \frac{1}{\widetilde{\beta}}\Bigg(\underbrace{(\widetilde{R}\beta)^{\sigma}a_0^i - ((\widetilde{R}\beta)^{\sigma} + \underline{r} + \phi)k_1^i - \varphi Y}_{\text{Intertemporal substitution}} + \underbrace{\frac{1}{2}\alpha^i\sigma^2_{c_2^i}}_{\text{Precautionary savings}}\Bigg).$$

Lemma 5 is a generalized counterpart to Proposition 1 in Angeletos and Calvet (2006) derived under a baseline CARA specification. The two-period structure leads to an intertemporal substitution effect due to the risky asset holdings and a precautionary savings effect that arises from uncertainty about the realization of second period consumption c_2^i. In this economy, the effects of a change in the shape of the wealth distribution on output can be characterized by means of sufficient statistics.

Corollary 1 (EFFICIENCY AND INEQUALITY). *The second period output in this economy is given by $Y = \mathbb{E}[b_1(\vartheta, a_0)]A + (\mu\phi + (1-\mu)A)\mathbb{E}[k_1(\vartheta, a_0)]$. The effect of a change in inequality on wealth-normalized output is*

$$\frac{\partial\widetilde{Y}}{\partial\eta} = A \cdot Cov\left(mps^i, \frac{da_0^i}{\mathbb{E}[a_0]}\right) + \mu(\phi - A)\cdot Cov\left(mpr^i \times mps^i, \frac{da_0^i}{\mathbb{E}[a_0]}\right),$$

where $mps^i \equiv \frac{\partial a_1^i}{\partial a_0^i}$ and $mpr^i \equiv \left(\frac{\partial k_1^i}{\partial a_1^i}\right)/\left(\frac{\partial a_1^i}{\partial a_0^i}\right)$ are the marginal propensity to save and to take risk.

Corollary 1 characterizes the overall effect of a wealth inequality change on aggregate efficiency into sufficient statistics in an economy with capital accumulation. In such an economy, type and scale dependence determine the extent to which agents invest in risky assets, captured by the distribution of mpr^i, and also how much they accumulate wealth, captured by the distribution of mps^i. As such, both covariance terms depend on γ, ρ and η in the aggregate. The quantitative setting in Section 3 also incorporates an explicit saving decision into a dynamic framework.

2.4.2 Extensions

Other sources of *scale* dependence Appendix OA 1.3.1 introduces an entrepreneurship type of model along the lines of Cagetti and De Nardi (2006), Guvenen et al. (2019) or Brüggemann (2021). The model is shown to map into the representation of equation (9). In this setting, we show that wealth-normalized output depends negatively on wealth inequality due to decreasing returns to scale on private equity investments (negative wealth-dependence $\gamma < 1$), but positively with the selection of entrepreneurs at the top of the distribution (positive type-dependence $\varrho > 0$). As stated before, such models are located in the bottom-right area of the inequality-efficiency diagram of Figure 1. Second, we consider the case of wealth-dependent borrowing constraint and show that it generates similar results as the one derived above under wealth-dependent risk-aversion.

Aggregate shocks Throughout the paper, we assumed that investment return risk is idiosyncratic following the findings of Bach et al. (2020) on private equity, which represents the largest share of wealth in the hands of the wealthy. We now check how our insights change under the assumption of aggregate production risk. Therefore, we assume that the productivity of innovative projects is stochastic and given by $z(\mu\phi + (1-\mu)A)$ with $z \sim \mathcal{N}(1, \sigma_z^2)$ an aggregate shock. In this case, a *growth – variance trade-off* arises as an increase in wealth inequality does not only affect expected growth, but also its volatility. A social planner seeking to redistribute wealth has an additional incentive to stabilize the wage rate, pushing towards less inequality when higher inequality is linked to higher aggregate risky investments. Under the special case of Section 2.2, this creates a positive link between inequality and wealth-normalized output volatility $\sigma^2_{\tilde{Y}}(\eta)$ if a certain model economy falls into the growth-enhancing region regarding type and scale dependence. Overall, the main insights and trade-offs of interest remain valid under this assumption.

3 A Dynamic Quantitative Model with Investment Heterogeneity

Section 2 derived an analytical representation of the link between wealth inequality, aggregate output and welfare. We isolated four key parameters: (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking to wealth, (iii) the sorting of types along the wealth distribution, and (iv) the extent to which returns to wealth reflects higher investment productivity. Focusing on those elements, we now build a quantitative model in which the wealth distribution arises endogenously, and study the distinct effects of type and scale dependence for wealth taxation.

3.1 Environment

The distribution of wealth arises endogenously from two empirically relevant features: heterogeneity in labor productivity as in a standard incomplete markets model (Aiyagari, 1994) and heterogeneity in capital investment and associated returns. While Benhabib et al. (2011, 2019) and Hubmer et al. (2020) show that the latter is key to generate the right tail of the wealth distribution, we study the role of the sources of this heterogeneity, type and/or scale dependence, for wealth accumulation and redistribution. We also introduce a life-cycle structure with endogenous labor supply which has been previously demonstrated in the literature to have potential for generating positive capital taxes (Conesa et al., 2009). Apart from those elements, the rest of the model is kept deliberately parsimonious.

3.1.1 Demographics, preferences and endowments

Time is discrete. Households derive a per period flow of utility $u(c_t^i, \ell_t^i)$ from consumption c_t^i and labor supply ℓ_t^i. At the beginning of each period agents differ in their wealth a_t^i, their age bracket j_t^i, their permanent component of labor productivity h_t^i, and their innate risk taking or investment skill type ϑ_t^i. They discount future periods at rate $\beta \in (0,1)$ and die with probability $d_j^i \equiv d(j_t^i)$. The expected life-time utility is given by

$$\mathcal{W}^i = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(1 - d_j^i\right)^t u(c_t^i, \ell_t^i) \right] . \tag{12}$$

Unless necessary, we drop the time and households indexes. Our model incorporates stochastic aging to capture the dynamics of income and wealth accumulation over the life-cycle. We thus assume that agents live through a discrete number of stages, i.e. $j \in \mathcal{J} \equiv [1, \ldots, J]$. From stage 1 to $J-1$ households participate in the labor market. Stage J comprises households beyond retirement. The probability of switching between age brackets j and $j+1$ is denoted by $\pi_j(j+1|j)$. Upon death, a household is replaced by a newborn household who inherits their wealth. There are no annuity markets such that households leave unintended bequests.

Individuals who work earn pre-tax labor income defined by $wy\mathcal{H}(h)\zeta_j\ell$, where w is the equilibrium wage rate, $y \sim F_y(y)$ and $\mathcal{H}(h)$ denote respectively the transitory and the persistent labor productivity components, while ζ_j is the age component of earnings. The evolution of the persistent component follows a first order Markov chain with transition probability $\pi_h(h'|h)$. At retirement, we assume that the working ability h stays constant over time, such that pensions are given by $w\mathcal{H}(h)\zeta_J$, with ζ_J defining the replacement rate. Upon death, a newborn imperfectly inherits the persistent component of her parents. With probability p_h she draws her parent's persistent labor productivity, and with probability $(1 - p_h)$ she draws her productivity from the invariant distribution $F_h(h)$ generated by $\pi_h(h'|h)$.

The risk-taking type $\vartheta \in \{\vartheta_1, ..., \vartheta_S\} \in \Theta$ follows a Markov chain with transition probability $\pi_\vartheta(\vartheta'|\vartheta)$. A newborn draws her parent's risk-taking type with probability p_ϑ and from the invariant distribution $F_\vartheta(\vartheta)$ otherwise.

Households are heterogeneous in their capital investments. They split their savings into safe and risky assets. An agent with risk-taking type ϑ and wealth a invests a share $\omega(a, \vartheta)$ in the risky asset. For the sake of clarity, our portfolio specification should be understood as a reduced form of a more elaborated portfolio choice. Let r_F and r_R, with $r_R > r_F$ be the safe and risky net returns determined in equilibrium, respectively. The pre-tax return on total investment is given by

$$r(a, \vartheta, \kappa) = \underline{r} + r_F \cdot (1 - \omega(a, \vartheta)) + (r_R \kappa) \cdot \omega(a, \vartheta) \, , \tag{13}$$

where $\underline{r}$ is an aggregate return component and $\kappa \sim F_\kappa(\kappa)$ an idiosyncratic element of *luck*. The variance of returns is therefore given by $\sigma_r^2 = (r_R\omega(a,\vartheta))^2\sigma_\kappa^2$ and implies that households with higher equity shares experience higher portfolio risk. Such a feature is supported in the PSID, and documented by Bach, Calvet and Sodini (2020) and Fagereng, Guiso, Malacrino and Pistaferri (2020). From equation (13), it is clear that returns are correlated over time through wealth itself, and through the process governing the evolution of investment skill-type ϑ.

Agents optimally choose their saving a', labor supply ℓ, consumption c and cannot borrow. Their recursive program is

$$v(a,\vartheta,h,j) = \mathbb{E}_{\kappa,y}\left\{\max_{c>0,\,a'\geq 0,\,\ell\geq 0}\left\{u(c,\ell)+\beta(1-d_j)\mathbb{E}_{j',h',\vartheta'|j,h,\vartheta}\left[v(a',\vartheta',h',j')\right]\right\}\right\} \tag{14}$$

$$\text{s.t.}\quad c+t_c(c)+a' = \mathcal{Y}^{inc}-t_w(\mathcal{Y}^{inc})+r(a,\vartheta,\kappa)a-t_r(r(a,\vartheta,\kappa)a)+a-t_a(a)\ , \tag{15}$$

$$\mathcal{Y}^{inc} = w\mathcal{H}(h)\left(\mathbb{1}_{\{j<J\}}\zeta_j y\ell+\mathbb{1}_{\{j=J\}}\zeta_J\right)\ , \tag{16}$$

where the functions $t_r(\cdot)$, $t_w(\cdot)$, $t_c(\cdot)$ and $t_a(\cdot)$ are taxes on capital income, labor income, consumption and wealth. Upon death, bequests are taxed such that $a^{child} = a^{parents} - t_b(a^{parents})$.[21]

3.1.2 Production, government, and equilibrium

Production The production sector is similar to the one in section 2. An intermediate producer operates at no cost a continuum of projects in sectors $s \in \{N, I\}$. Each project uses assets supplied by a household with wealth holdings a and skill type ϑ to produce x intermediate goods with technology

$$x^N(a,\vartheta) = A[(1-\omega(a,\vartheta))a]^{\nu_N}\ ,\qquad x^I(a,\vartheta) = (\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I}, \tag{18}$$

where $\phi > A$ holds and $(\nu_N,\nu_I)\in\mathbb{R}_+^2$ are returns to scale on the technologies. Similar to section 2, μ is a wedge capturing the extent to which returns on risky investments reflect the associated capital productivity. The intermediate producer sells intermediate goods to a final good producer at price $p^s(a,\vartheta)$ and obtains revenues $\Pi(a,\vartheta) = \sum_s p^s\cdot x^s\,(a,\vartheta)$, which are redistributed to investors. Recall that intermediate producers do not face any risk, however, investors are exposed to the investment shock κ.

A competitive final good producer uses labor L and intermediate goods $X = \sum_s\left(\int_i x_i^s di\right)$ to produce with technology $Y = F\,(X,L)$, where $F(\cdot)$ satisfies the Inada conditions. Profit maximization, i.e. $\max_{x_i^s,L}\, Y - \sum_s\int_j\,(p_i^s+\delta)x_i^s\,di - wL$, yields the following set of prices: $p_i^s = \frac{\partial F(X,L)}{\partial X}\frac{\partial X}{\partial x_i^s} - \delta$,

[21] The outer expectation comes from the fact that y and κ are *iid*. An alternative way to write this value function is

$$v(a,\vartheta,h,j,\kappa,y) = \max_{c>0,\,a'\geq 0,\ell\geq 0}\left\{u(c,1-\ell)+\beta(1-d_j)\mathbb{E}_{j',\kappa',y',h',\vartheta'|j,h,\vartheta}\left[v(a',\vartheta',h',j',\kappa',y')\right]\right\}. \tag{17}$$

and $w = \frac{\partial F(X,L)}{\partial L}$, where $\delta \in (0,1)$ is the depreciation rate. As intermediate goods are perfect substitutes, it follows that $p_i^s = p \; \forall i, s$.

Given the intermediate goods equations (18), the return wedge and the profit maximization, the returns to wealth to safe and risky asset investments are given by

$$r_F := \frac{px^N(a,\vartheta)}{(1-\omega(a,\vartheta))a} = MPK_F = pA[(1-\omega(a,\vartheta))a]^{\nu_N - 1}, \tag{19}$$

$$r_R := \frac{px^I(a,\vartheta)}{\omega(a,\vartheta)a} = p\phi[\omega(a,\vartheta)a]^{\nu_I - 1} \geq MPK_R = p(\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I - 1}, \tag{20}$$

where $\mu < 1$ implies $r_R > MPK_R$. This thus describes the case in which risky returns to wealth do not only reflect investment productivity but also some form of rent-extraction, for example.

Government The government finances an exogenous expenditure level G as well as social security retirement pensions. It raises total revenues from consumption, capital income, bequest, wealth, and labor income taxes. Consumption, capital income and labor income are subject to a linear tax, respectively $t_c(\mathrm{x}) = \mathrm{x}\tau_c$, $t_r(\mathrm{x}) = \mathrm{x}\tau_r$, and $t_w(\mathrm{x}) = \mathrm{x}\tau_w$. There is no wealth tax in the baseline economy, i.e. $t_a(\mathrm{x}) = 0$. We will however study the case of a progressive wealth tax in section 6. The bequest tax is also linear, $t_b(\mathrm{x}) = \tau_b(\mathrm{x} - t_a(\mathrm{x}))$, where we assume that the wealth tax is paid first. Consequently, the government budget is given by

$$\begin{aligned} G + \int_{(a,\vartheta,h)} w\mathcal{H}(h)\zeta_J \, d\mathcal{G}(a,\vartheta,h,j=J) = \int_{(a,\vartheta,h,j)} \Big(& \tau_w \int_y \mathbb{1}_{\{j<J\}} w\ell\mathcal{H}(h)y\zeta_j \, dF_y(y) + \\ & + \tau_r \int_\kappa r(a,\vartheta,\kappa)a \, dF_\kappa(\kappa) + \tau_c c(a,\vartheta,h,j) \\ & + \tau_a(a) + \tau_b d_j(a - t_a(a)) \Big) d\mathcal{G}(a,\vartheta,h,j) \,. \end{aligned} \tag{21}$$

3.2 Equilibrium

In each period t, the aggregate state of the economy is described by the joint measure $\mathcal{G}_t$ over asset positions, labor productivity, investment skill-type and age.

Definition 2. *Denote the state space by $\boldsymbol{s} = (a,\vartheta,h,j) \in \mathbb{S} \equiv \mathbb{R}_+ \times \Theta \times \mathbb{H} \times \mathcal{J}$. A steady-state equilibrium of this economy is a vector of quantities $\{Y, X, L\}$, a set of policy functions $\{c(\boldsymbol{s}), a'(\boldsymbol{s}), \ell(\boldsymbol{s})\}$, a set of prices $\{p, w, \underline{r}\}$, a set of tax functions $\{t_w, t_r, t_c, t_b, t_a\}$, and a probability distribution of households $\mathcal{G}$ defined over $\mathbb{S}$, such that*

(1) The representative final producer maximizes profits, i.e. $\max_{\{X,L\}} F(X,L) - (p+\delta)X - wL$, *where p and w are given by their respective marginal products.*

(2) Given prices, households solve the stationary version of their decision problem (14)*, giving rise to an invariant distribution $\mathcal{G}(\boldsymbol{s})$.*[22]

[22]At each state (a,ϑ,h,j), there is a continuum of individuals experiencing the *iid* shocks y and κ. The stationary

(3) The government budget constraint (21) is satisfied.

(4) Labor and intermediate goods markets clear, i.e.

$$L = \int_y \int_{(a,\vartheta,h,j)} \mathbb{1}_{\{j<J\}} w\ell\mathcal{H}(h)y\zeta_j \, d\mathcal{G}(a,\vartheta,h,j) \, dF_y(y) \ , \tag{22}$$

$$X = \int_{(a,\vartheta,h,j)} \Big(A[(1-\omega(a,\vartheta))a]^{\nu_N} + (\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I} \Big) \, d\mathcal{G}(a,\vartheta,h,j) \ . \tag{23}$$

(5) The total capital product distributed by the intermediate producer for each project $\Pi(a,\vartheta)$ is consistent with the total capital income received by households, i.e.

$$pX = \int_\kappa \int_{(a,\vartheta,h,j)} \Big(\underline{r} + r_F(a,\vartheta)(1-\omega(a,\vartheta)) + r_R(a,\vartheta)\kappa\omega(a,\vartheta) \Big) a \, d\mathcal{G}(a,\vartheta,h,j) \, dF_\kappa(\kappa). \tag{24}$$

Finally, by Walras Law, the good market clears. Moreover, our measure of total wealth in this economy is given by $K = \int_{(a,\vartheta,h,j)} a \, d\mathcal{G}(a,\vartheta,h,j)$.

Condition (23) states that the total efficiency units of capital used in the production sector must correspond to the total capital supplied by households given their investment choice between risky and safe assets. When $\mu < 1$, each unit of risky investment yields returns to wealth higher than its corresponding marginal product. As such, the equilibrium base return $\underline{r}$ must be negative to satisfy condition (24). Therefore, besides X/L that pins down p and w, the aggregate return component $\underline{r}$ in equation (13) is the second object that adjusts in equilibrium.

3.2.1 Numerical solution

The model admits no analytical solution. We solve it numerically using a version of the endogenous grid method (Carroll, 2006). Appendix C.2 describes the algorithm. Under certain calibrations, e.g. the one of the benchmark economy, the model induces a Pareto distribution of wealth at the top. As a consequence, the support of the stationary distribution of wealth is unbounded from above. To circumvent this issue in practice, we use a large value for the upper bound on wealth in our numerical implementation.[23]

distribution is obtained using non-stochastic simulations.

[23] To check the size of the error implied by this truncation, we estimate the Pareto tail, $\widehat{\eta}_\xi$, of the model-generated wealth distribution at the top above the a_ξ wealth threshold, which is chosen large enough such that the upper part of the distribution is indeed Pareto. The details of the estimation procedure are similar to the one implemented in Appendix B.1. In the truncated model, for $a \geq a_\xi$, there is a mass $\mathcal{G}_\xi$ of households at the top with total wealth K_ξ who thus hold a fraction $\theta_K = K_\xi/K$ of total wealth in the economy. We then reconstruct a Pareto distribution with shape $\widehat{\eta}_\xi$ and scale a_ξ, and compute the implied fraction $\widehat{\theta}_K$ from this theoretical distribution. We increase the upper bound of the grid on wealth until $|\widehat{\theta}_K - \theta_K|$ is negligible.

4 Taking the Model to the Data

Before studying wealth taxation, we first choose a parameterization that allows the model to closely replicate key features of the US economy. We begin with a discussion of the data used.

4.1 Capital Heterogeneity: A First Glance into the Data

We use the SCF and the PSID micro datasets. Our concept of wealth is *net worth*, which is defined as the market value of all assets minus total liabilities. Assets comprise riskless assets (deposits, savings, cash), direct and indirectly (mutual funds, individual retirement accounts) held public equities, net private equity business investments, primary and secondary residences, and other non-financial assets. Liabilities comprise student loans, mortgages, consumer credits and other loans. We restrict the sample population to those aged 20–70, essentially to focus on decision makers. We define private and public equity as productive risky assets while assets such as riskless assets, residential properties and other non-financial assets are considered as safe productive assets.[24] This classification is in line with the literature; for example, Cagetti and De Nardi (2006) separate private equity assets from other investments into corporate firms, while Kaplan et al. (2018) consider only equity and commercial or business real estate as productive assets.

Our SCF waves span from 1989 to 2019 and include detailed information on households' portfolio composition, comprising a number of very wealth-rich households. When computing moments related to wealth inequality, we will sometimes refer to the *adjusted SCF*. The adjustments are made using the method of Vermeulen (2016) for *all* periods. First, we correct for underreporting of assets by adjusting survey estimates of real assets, financial assets, and liabilities such that they align with aggregate national balance sheets. Second, we adjust for under-representation at the very top by merging the SCF with households from the Forbes World's Billionaires lists and extrapolate wealth shares based on estimating a Pareto Law. Appendix B.1 details the procedure.

Additionally, we use PSID waves from 1998 to 2018 to compute empirical moments of investment returns and extract complementary information on investors. The PSID constitutes a large and representative biennial survey with a long panel dimension that contains information by broad asset classes on capital income and costs, asset prices, inflows and outflows.[25]

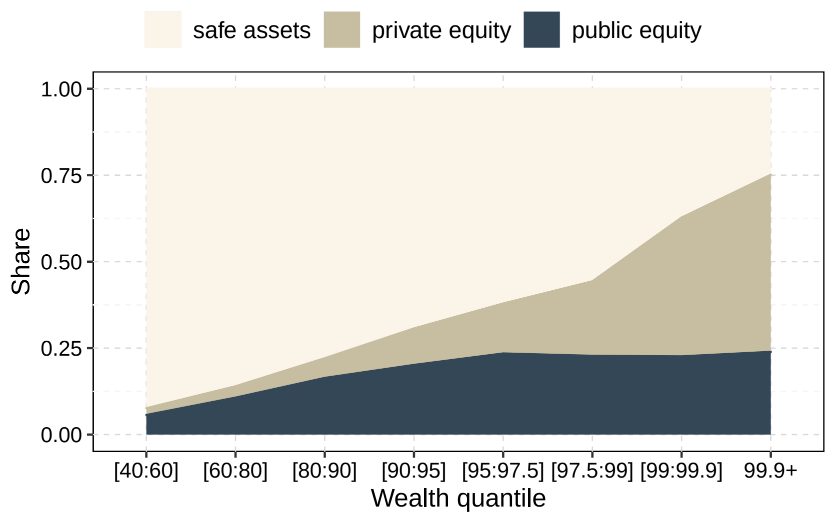
Portfolio composition. In Figure 3 we first use the SCF to get a full *cross-sectional* picture of the average household's portfolio composition across the US wealth distribution. As already documented in the literature, private and public equity investments correlate strongly and positively

[24]Our motivation for classifying housing in safe assets comes from the fact that returns from housing assets are less volatile than other assets. Moreover, their volatility relates mostly to differential mortgage interest rates.

[25]See Pfeffer et al. (2016) for an excellent comparison between both surveys and Flavin and Yamashita (2002) for a discussion about returns estimated from the PSID. Appendix B.2 provides further details. A drawback of the PSID is the presence of only a small number of households at the very top (within the top 1%) of the wealth distribution.

with wealth in the cross-section. The top 1% in the US hold on average 65% of wealth in risky equity, while the corresponding share for the median household is 7%.[26]. As it can be seen, the noticeable increase in risky equity share at the very top of the wealth distribution (above the 95^{th} percentile) is mostly driven by assets held in private equity business investments.

Figure 3. Average portfolio share of gross assets by wealth percentile.



Source: adjusted SCF from 1989 to 2019, averaged from different SCF imputations.

Returns to wealth. We measure returns to wealth in the US using the PSID. We compute returns for each broad asset class $l \in \{riskfree, home, secondary, priv, public, other\}$, where $riskfree$ assets bundle savings, bonds and checking accounts, and *home* denotes assets linked to the primary residence. As stated before, risky assets comprise public equities and non-financial private equity business investments. The return on asset class l for household i in year t is given by

$$r_{i,l,t} = \frac{R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t}}{a^g_{i,l,t-1} + F_{i,l,t}/2}\,, \tag{25}$$

where $a^g_{i,l,t-1}$ is the beginning-of-period amount of asset class l held and $F_{i,l,t}$ are inflows minus outflows (i.e. net investment), that we divide by two assuming that they occur in mid-year.[27] The values $R^K_{i,l,t}$ and $R^I_{i,l,t}$ and $R^D_{i,l,t}$ correspond respectively to capital gains, asset income such as dividends, interests and other payments and to the cost of debts (if any). Returns to total net worth are similarly computed as $r_{i,net\,worth,t} = \frac{\sum_l R^K_{i,l,t}+R^I_{i,l,t}-R^D_{i,l,t}}{\sum_l (a^g_{i,l,t-1}+F_{i,l,t}/2)}$. Finally, nominal returns are converted to real returns using the consumer price index for each year. The complete and detailed procedure regarding the construction of returns is provided in Appendix B.2.

[26]These numbers are comparable to estimates from detailed Swedish administrative data as in Bach et al. (2020), who use a similar definition of risky assets

[27]This is mostly due to reduce the bias due the acquisition and sale that generate large returns (i.e. division by zero).

Table 1 provides selected descriptive statistics on the returns to net worth (before-tax), to riskless asset, to public equity and to private equity. Notably, returns to private equity display the highest expected returns (15.6%) with substantial heterogeneity and skewness to the right. To a lower extend, public equity generates also substantial returns (5.8%). Despite the absence of very wealthy households in the PSID, our results are comparable to estimates in Fagereng et al. (2020) in Norway and Bach et al. (2020) in Sweden using administrative data. As a direct comparison to US estimates, Xavier (2020) evaluates, using cross-sectional information from the SCF, that aggregate returns are 13.6% for private equity, 6.4% for public equity, and between 0.4%-2.1% on the different safe assets.[28] A noticeable difference, however, is that our aggregate estimate of returns to net worth (before-tax) is 3.3%, which is substantially lower than the one evaluated by Xavier (2020) (6.8%), but closer to the ones estimated using a quantitative structural model of inequality by Benhabib et al. (2019) (3.1%) and to the empirical estimates in Fagereng et al. (2020) for Norway (3.8%). In comparison to Sweden, Bach et al. (2020) find a median return to net worth of 4.5% with a standard deviation of 13% per year. We attribute this discrepancy to the fact that the PSID does not account for very wealthy households and to using different methodologies.

Table 1. Returns to wealth in the PSID.

Wealth component	Descriptive Statistics					
	Mean	St.Dev.	Skewness	Kurtosis	20^{th} perc.	80^{th} perc.
Net worth (before-tax)	0.033	0.158	0.897	6.243	-0.035	0.089
Private equity	0.156	0.614	2.071	10.967	-0.225	0.500
Public equity	0.058	0.417	-0.122	0.085	-0.248	0.385
Riskless assets	0.004	0.009	3.234	10.267	0.000	0.003

Note: we apply a trimming of 0.5% at the top and the bottom for each asset class.

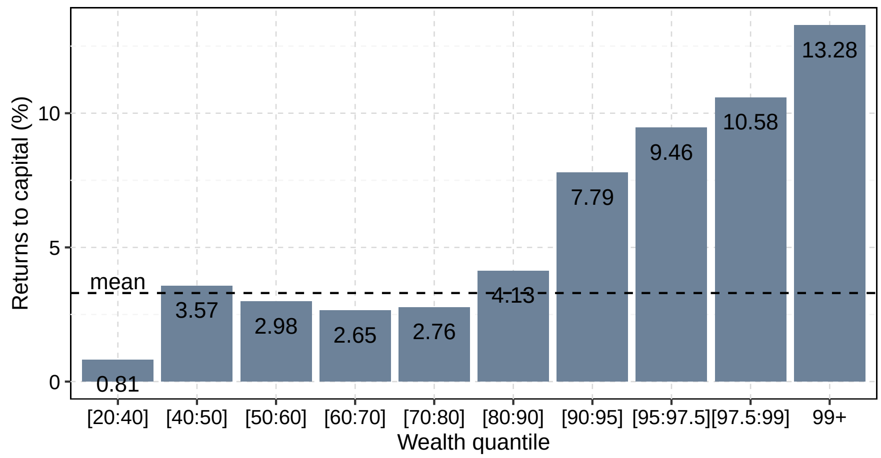
Figure 4 documents the before-tax returns to wealth by wealth quantile in the PSID. Returns to wealth positively correlate with wealth in the cross-section, which is likely to be driven by heterogeneity in household's portfolio composition, as documented above.[29] These findings are consistent with existing work establishing a positive correlation between private equity ownership and wealth (Quadrini, 2000; Cagetti and De Nardi, 2006). Of course, those numbers are not informative on whether the correlation is driven by type or scale dependence. In practice, there

[28]Her methodology is very different from ours. We use the panel dimension of the PSID to compute returns of a given household over time, while she computes returns from the SCF waves using cross-sectional information on capital income and stocks, averaged by wealth percentile. A drawback of her analysis is sample selection, as individuals may in principle move in and out of a given wealth percentile over time, as shown in Gomez et al. (2018). Contrary, a drawback of our analysis is the under-representation of very wealthy households in the PSID.

[29]This inference is shared by Bach et al. (2020) while Fagereng et al. (2020) find that there remains substantial heterogeneity within a broad asset class, which may reflect an important role for heterogeneity in skills. In fact, conditional on a broad asset class, it is difficult to disentangle whether higher returns are due to specific skills or due to higher risk-taking. For private equity investments, the latter may arise due to diversification motives (Penciakova, 2018) or due to the interaction between business risk-taking and borrowing limits (Robinson, 2012). In Appendix B.2, we decompose the *cross-sectional* relation and find that this correlation is not observed *within* asset classes.

is no obvious way to disentangle the two as both channels are likely to drive the observed cross-sectional relationship, as we will demonstrate subsequently.

Figure 4. Average return on wealth by gross wealth quantile.



To sum up, the increasing share of risky assets at the top of the wealth distribution is substantially driven by private and public equity holdings displaying the highest expected returns. Among other possible determinants, the positive correlation between returns to wealth and wealth *itself* is thus likely to be driven by differing portfolio composition.

4.1.1 Type and scale dependence in capital investments in the US economy

Exploiting the panel dimension of the PSID by controlling for individual characteristics, Hurst and Lusardi (2004) find evidence for scale dependence in the propensity to select into private equity business ownership among the top 5% wealthiest households. Their estimates show that the average probability is flat at a rate of 3 percent for the bottom 80%. It increases to 4 percent for the 95^{th} percentile and reaches 7% for the 98^{th} percentile.

Using information on returns to capital endowments of US universities, Piketty (2018) (Chapter 12) finds that returns substantially increase with wealth and argue that this may arise from economies of scale in portfolio management.[30] However, US universities may not be representative of the US population as their investment strategies may substantially differ.

Bach et al. (2020) use Swedish administrative panel data and test for scale dependence in returns to wealth. According to them, even within a sample of twins and controlling for twin-pair fixed effects, there is strong evidence of scale dependence, especially at the top of the wealth distribution (Table 9, p. 2738). They argue that scale dependence is likely driven by changes in the individual portfolio composition, e.g. due to returns to scale on management costs or DRRA behavior. Controlling for individual and year fixed effects, Fagereng et al. (2020) use a Norwegian

[30] In Piketty's words, "*The most obvious one is that a person with 10 million euros rather than 100,000, or 1 billion euros rather than 10 million, has greater means to employ wealth management consultants and financial advisors.*".

administrative panel on wealth tax records and regress the average return to wealth on the individual's wealth percentile at the beginning of the period. Both wealth (scale) and individual fixed effects are found to be statistically significant. Their estimates imply that scale dependence alone explains 48% of the 18 percentage point return difference between the 10^{th} and 90^{th} net worth percentiles.

Finally, Robinson (2012) shows that wealthier private equity business owners tend to start relatively riskier business investments, with higher expected profitability. Penciakova (2018) confirms a similar pattern with data on US firms using the Census Bureau's Longitudinal Business Database, patenting data, and Compustat. She documents that private equity investors who diversify start riskier additional private equity investments. As we will show, this diversification among private equity owners occurs mostly among the wealthiest households.

All in all, this leads us to conclude that both type and scale dependencies are likely to drive the observed correlation between returns to wealth and wealth.

4.2 Functional Forms and Calibration

The benchmark model aims to capture features of the SCF and PSID described above. Apart from this, the model is parameterized to account for realistic government policy, demographics, labor income process and production technology.

We map the stationary equilibrium of the quantitative model to US data in two steps. We first fix some parameters based on model-exogenous information. In a second step, we calibrate the remaining parameters endogenously by numerically simulating the model to match particular empirical moments. Table 2 summarizes all parameters.

4.2.1 Exogenously set common parameters

Preferences and technology The model is calibrated to the US economy and the period is one year. Preferences over consumption c and labor supply ℓ are represented by a standard time-separable utility function of the form

$$u(c,\ell) = \frac{c^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \chi \frac{\ell^{1+\frac{1}{\lambda}}}{1+\frac{1}{\lambda}} \,. \tag{26}$$

We set the IES $\sigma = 0.5$, and the Frisch labor elasticity $\lambda = 0.6$ following Brüggemann (2021) and Kindermann and Krueger (2014). The disutility cost χ is chosen so that households spend, on average, 1/3 of disposable time on market work. The discount factor β matches a capital-output ratio $\frac{K}{Y}$ of 2.6, which is consistent with Kitao (2008).

Demographics We model four age brackets. The first three brackets span from age 20 to age 65, each with a length of 15 years. The probability of switching from age bracket j to the next one

Table 2. Calibrated parameters.

PARAMETER	SYMBOL	VALUE	SOURCE/TARGET
Demographics and Preferences			
Survival probability	d_j	in text	Data (US social security statistics) [a]
Discount factor	β	0.925	Capital-output ratio of 2.6
Risk aversion	σ^{-1}	2.0	Conesa et al. (2009)
Frisch elasticity	λ	0.6	Kindermann and Krueger (2014)
Disutility of labor	χ	18	1/3 of time on market work
Labor productivity process			
Age component	ζ_j	in text	Guvenen et al. (2021)
Permanent component	$\{\rho_h, \sigma_h\}$	$\{0.95, 0.2\}$	Storesletten et al. (2004)
Transitory component	$\{\sigma_y\}$	0.15	Heathcote et al. (2010)
Pareto tail	$\{\eta_h, q_h\}$	$\{1.9, 0.9\}$	Piketty and Saez (2003)
Intergenerational corr. h	p_h	0.35	Chetty et al. (2014)
Low income state	y_0	0.08	12% households with zero wealth (SCF)
Investment			
Intergenerational corr. ϑ	p_ϑ	0.15	Fagereng et al. (2020)
Risky share param. – scale	in text	in text	Shape of portfolio distribution (SCF)
Risky share param. – type	$\underline{\omega}$	0.4	Conditional risky share (SCF)
Transition between types	in text	in text	Data (PSID)
Excess wealth return	ϕ	6.2	Top 1% wealth share of 0.36
Technology			
Labor share	$1-\alpha$	0.67	Data
Depreciation	δ	0.045	Assumption
Safe technology constant	A	1.0	Normalization
Returns to scale	$\{\nu_N, \nu_I\}$	1.0	Assumption
Government policy			
Consumption tax	τ_c	0.05	Conesa et al. (2009)
Labor income tax	τ_w	0.225	Guvenen et al. (2019)
Capital income tax	τ_r	0.25	Guvenen et al. (2019)
Bequest tax	τ_b	0.4	Data, statutory tax rate

[a] See https://www.ssa.gov/oact/STATS/table4c6.html.

$j+1$ equals accordingly $\frac{1}{15}$ for agents in the first three age brackets, while retired households stay in the terminal bracket J until death. The death probability d_j for each age bracket j is taken from the US social security statistics and is reported in Table 3.

Technology We specify $F(X, L) = X^\alpha L^{1-\alpha}$ with $\alpha = 0.33$ and set the depreciation rate to 4.5%. We normalize $A = 1$. In the baseline, we set $\mu = 1$ such that $r_F = MPK_r$, and consider later cases with $\mu < 1$ when analyzing the aggregate effects of wealth taxation.

As already stressed above, micro returns to scale on investment (ν_N, ν_I) constitute another important form of scale dependence on returns. Existing data from Bach et al. (2020) find no evidence on the presence of increasing or decreasing returns on risky investments, especially on private eq-

Table 3. Life cycle earning profile, mortality and transition probabilities.

	ζ_j	d_j	$\pi_j(j+1\|j)$ conditional on survival			
			$[20,35)$	$[35,50)$	$[50,65)$	≥ 65
$[20,35)$	0.81	0.15%	0.933	0.067	0.0	0.0
$[35,50)$	1.00	0.4%	0.0	0.933	0.067	0.0
$[50,65)$	1.35	1.1%	0.0	0.0	0.933	0.067
≥ 65	0.40	9.9%	0.0	0.0	0.0	1.0

uity returns.[31] We thus set $\nu_N = \nu_I = 1$. However, it contrasts with the variety of entrepreneurship models assuming a DRS technology on risky private equity businesses investment (Cagetti and De Nardi, 2006; Brüggemann, 2021; Guvenen et al., 2019). Therefore, as CRS is not a standard assumption, we will also investigate the case where $\nu_I < 1$.

Government policies The government levies consumption, labor income, capital income and bequest taxes which approximately equal the current rates of the US economy. Labor income and capital income tax rates are set to $\tau_w = 22.5\%$ and $\tau_r = 25\%$, which is consistent with Guvenen et al. (2019). The bequest tax rate is fixed to $\tau_b = 40\%$ according to the corresponding statutory tax rates. Finally, the consumption tax is set to 5%, which is consistent with the value used in Conesa et al. (2009). Given the tax rates in place, the share of total government expenditure to GDP is equal to 0.24 in the stationary equilibrium, respectively 0.17 without social security payments.

Labor income process As stated above, a household's labor productivity depends on three components: an age-dependent component ζ_j, a persistent component h and a transitory component y. The natural logarithm of the individual hourly wage of a household writes

$$log(w) + log(\mathcal{H}(h)) + log(y) + log(\zeta_j) \ . \tag{27}$$

The life-cycle average earning profile ζ_j is taken from Guvenen et al. (2021) (Supplementary Appendix, Figure C.36). Table 3 provides the parameter values used throughout this paper for each age bracket, including the transition probability matrix across age brackets.

The processes for labor productivity aim to generate a realistic earning distribution and contribute to the overall wealth inequality. We follow Hubmer et al. (2020) and improve the fit of the earnings distribution by assuming that the persistent component $\mathcal{H}(h)$ follows a lognormal AR(1) process with persistence ρ_h and variance σ_h^2. However, at the top of the income distribution, h is

[31] Furthermore, we use our PSID sample and estimate, fixing the broad asset class l, the effect of asset holdings $a_{i,l,t}$ on asset returns $r_{i,l,t}$ according to $\log(r_{i,l,t}) = \beta_l \log(a_{i,l,t}) + FE_i + FE_t + \epsilon_{i,l,t}$, where β_l reflects the returns to scale, while FE_i and FE_t stand for household and year fixed effects. If β_l does not statistically differ from zero, the CRS hypothesis on returns cannot be rejected. Our estimates suggest that there is no statistical evidence for either IRS or DRS, even among private equity business holdings.

drawn from a Pareto distribution with shape $\eta_h > 1$,

$$\mathcal{H}(h) = \begin{cases} e^h & \text{if} \quad F_h(h) \leq q_h \ , \\ F^{-1}_{Pareto(\eta^h)}\left(\frac{F_h(h)-q_h}{1-q_h}\right) & \text{otherwise} \ , \end{cases} \tag{28}$$

where $F_h(h)$ is the CDF of h and $F^{-1}_{Pareto(\eta^h)}(\cdot)$ the inverse CDF for a Pareto distribution with lower bound $F_h^{-1}(q_h)$ with $q_h \in [0,1]$. The persistent component is discretized into nine bins $h \in \mathbb{H} \equiv \{h_1, ..., h_9\}$. We set the threshold of the Pareto distribution to $q_h = 0.9$ and the shape to $\eta_h = 1.9$, consistent with 1990-2010 estimates for the US (Piketty and Saez, 2003). In line with Storesletten, Telmer and Yaron (2004), parameters of the persistent labor productivity component are set to $\rho_h = 0.95$ and $\sigma_h = 0.2$.[32] The correlation between parents' labor productivity with the one of their heir is $p_h = 0.35$, which is consistent with Chetty et al. (2014).

Following Heathcote et al. (2010), the transitory process follows a log-normal distribution, i.e. $y \sim \mathcal{LN}(0, \sigma_y^2)$, where $\sigma_y = 0.15$. The process y is discretized into three states using Gauss-Hermite quadratures. We further add a low income state y_0 that occurs with probability $\pi_y(y_0) = 7.5\%$ and reflects for instance involuntary unemployment or part-time work, independently of (y, h) and over time. We choose y_0 to generate a realistic fraction of individuals with zero wealth.

4.2.2 Calibrating capital heterogeneity and wealth returns

We now calibrate variables associated with returns to wealth. The standard deviation of risky returns σ_κ is set to 0.45; a value consistent with our PSID estimates for public and private equity (Table 1) and, for instance, Fagereng et al. (2020). The excess return parameter $\phi = 6.2$ generates a top 1% wealth share of 36% by controlling the dispersion of returns to wealth between households. As we will show later on, this value also produces a realistic distribution of returns to wealth.

To carefully pin down the type and scale dependence, one would need sufficiently detailed panel data on investment decisions to test for a statistical relation between portfolio composition, investment returns and wealth while controlling for household characteristics. Fagereng et al. (2020) and Bach et al. (2020) follow this strategy using administrative data, and find strong support that returns feature both type and scale dependence, acknowledging a crucial role for portfolio composition.[33] However, the results are conditioned by the statistical model used to estimate type and scale dependence, e.g. a linear model with fixed effects (type) and wealth percentiles (scale).

We pursue another strategy by recognizing that there are two common ways to generate scale

[32]For the sake of transparency, we reduce the computational burden by using a reduced transition matrix $\hat{\Pi}_h(h'|h)$ such that: $\hat{\Pi}_h(h'|h) = \begin{cases} \Pi_h(h'|h) & \text{if} \quad \Pi_h(h'|h) \geq \epsilon, \\ 0 & \text{otherwise.} \end{cases}$ with $\epsilon = 10e^{-6}$ and normalizing the matrix $\sum_{h'} \hat{\Pi}_h(h'|h) = 1$. This allows us to exploit the sparsity of the transition matrix.

[33]Moreover, notice that the role of private equity is substantial in Fagereng et al. (2020): *"All in all, heterogeneity in our most comprehensive measure of returns to wealth can be traced in the first place to heterogeneity in returns to private equity and the cost of debt and only partially to heterogeneity in returns to financial wealth"*.

dependence in models featuring private or public equity investments. On one hand, the extensive margin decision to invest might be wealth-dependent. This is the case in many occupational choice models in the presence of borrowing constraint (Cagetti and De Nardi, 2006; Brüggemann, 2021) or models with fixed participation cost (Fagereng et al., 2017). On the other hand, conditional on being an equity investor, there might be an intensive margin effect such that the portfolio share of investors is *itself* a function of wealth.

To capture both margins, we first assume that there are two investment skill-types, $\vartheta \in \{\vartheta_1 = 0, \vartheta_2 = 1\}$, i.e. those with investment skills, and those without, respectively. The probability of switching from an unskilled to a skilled investor type is a function of wealth, such that

$$\pi_\vartheta(\vartheta'|\vartheta, a) = \begin{bmatrix} 1 - \overline{\pi}_\vartheta - \lambda(a) & \overline{\pi}_\vartheta + \lambda(a) \\ \underline{\pi}_\vartheta & 1 - \underline{\pi}_\vartheta \end{bmatrix}, \tag{29}$$

where the components $\underline{\pi}_\vartheta$ and $\overline{\pi}_\vartheta$ capture switching probabilities that are unrelated to wealth, e.g. time-variations in skills or preferences. In the data, the fraction of public equity investors is substantially larger than the fraction of private equity holders. Moreover, the adjustment margin of portfolio allocation at the top of the wealth distribution comes mainly from private equity business investments (cf. Figure 3). As such, households investing in public equity and who do not invest in private equity are counted as investor only when they hold more than 50 percent of their wealth in public equity. This assumption means that 15% of households are investors. We set the exit probability $\underline{\pi}_\vartheta = 0.10$ consistent with the PSID and choose $\overline{\pi}_\vartheta = 0.018$ to match the fraction of investors. We calibrate $\lambda(a)$ using the following parametric form:

$$\lambda(a) = \min\{\lambda_1(\max\{a - \underline{a}_\lambda, 0\})^{\gamma_\lambda}, \lambda_2\}. \tag{30}$$

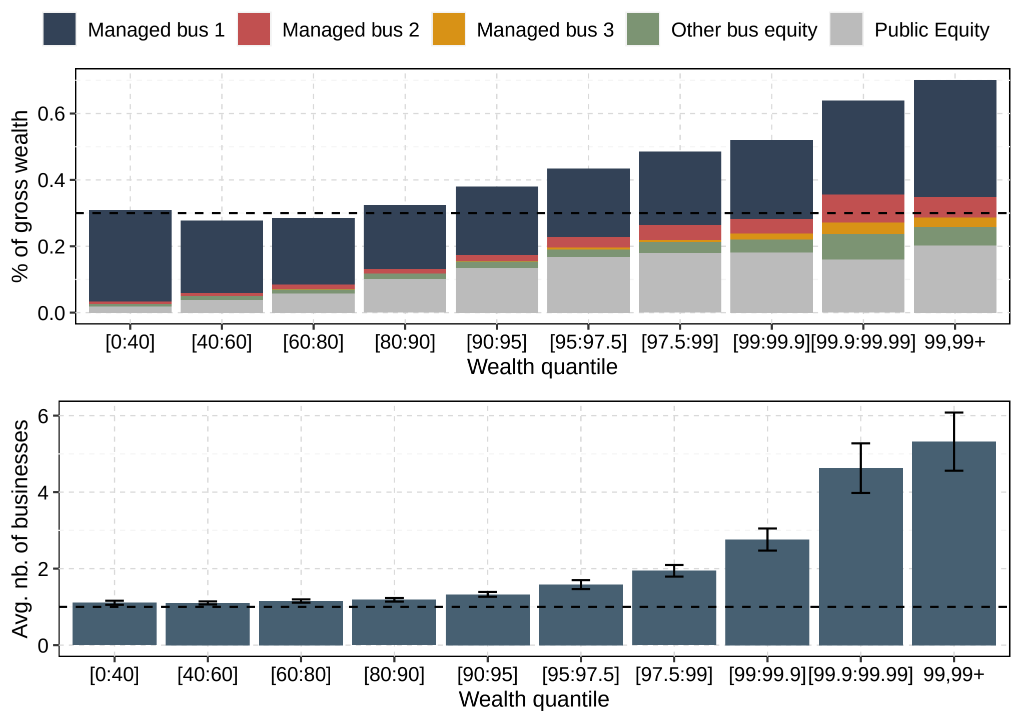
Using the results in Hurst and Lusardi (2004) derived from the PSID, parameters $\{\underline{a}_\lambda, \lambda_1, \lambda_2, \gamma_\lambda\}$ are chosen to generate the increasing probability to become an investor as a function of wealth, conditional on household's characteristics.[34] Consistent with the data, $\underline{a}_\lambda$ is set such that the probability to become an investor starts to be a function of wealth only above the wealth level corresponding to the 80^{th} wealth percentile. Below this wealth percentile, the participation rate is only generated through the process governing the ϑ-type, i.e. through $\overline{\pi}_\vartheta$. The level and the shape parameters of the transition probability with respect to wealth are endogenously set to $\lambda_1 = 0.071$ and $\gamma_\lambda = 0.30$ in order to replicate the average transition rate of 3.2% for households within the [95-97.5] wealth quantile, and of 6.1% for households within the [99-99.9] wealth quantile. A maximal value of $\lambda_2 = 0.045$ matches a transition rate of 7% for the top 1% wealthiest households.

We calibrate the intensive margin of risky investment as follows. In the SCF, conditioning on

[34] To save on space, we relegate the full empirical analysis of this observed relationship to Appendix B.3.

being private equity investor, the share of risky equity increases with wealth in the cross-section. One practical issue, however, is to distinguish whether the share of private equity held by those owners increases because of systematic capital gains or whether it is the result of net investments. Even in the latter case, it is difficult to disentangle whether private equity owners who obtain higher returns are more likely to end up at the top, or whether wealth itself induces households to undertake riskier investments. To circumvent those issues, we use the SCF and exploit detailed information on the biggest three household's private equity business investments and a bundle of the remaining ones, including their acquisition date and the share of wealth in each private equity.

Figure 5. Average share held in equity (top panel) and average number of private equity business investments (bottom panel), conditioning on investors.



Source: SCF (1989-2019). The dashed line represents the median.

The top panel of Figure 5 shows the average share of private equity investments per net worth quantile, decomposed in different business investments. The average share of private equity investment over total gross wealth appears to be strongly correlated with wealth, especially above the top 5%. This relation is driven by diversification at the top. From our standpoint, this pattern cannot be solely driven by type dependence. First, borrowing constraints are likely to prevent relatively wealth-poor households to invest in multiple businesses, limiting the concern regarding the possible reverse causality that an owner of multiple businesses is more likely to select over time at the top of the wealth distribution. Second, those additional businesses are generally newly founded; 85% of the second businesses were created in the past ten years relative to the survey date, and 67% were created within the past five years. As a comparison, 47% of the first main business of those private equity owners were created within the past ten years. Therefore, given that wealth accumulation is a slow process, it seems unlikely that multiple business owners at the top became rich due to multiple private equity investments. Instead, the decision to open additional private equity investments is likely, among other determinants, to be wealth-driven.[35]

We further elaborate on this point using two additional pieces of evidence. First, as shown in the bottom panel of Figure 5, the average *number* of private equity business investments substantially rises at the upper end of the wealth distribution. Again, looking at the timing of those additional businesses, the last acquired business is particularly recent relative to the main business. Finally, despite the lack of observations at the very top in the PSID, we use its panel dimension and confirm that investment in additional private equity business investments, conditional on being a business owner and controlling for individual characteristics, is statistically positively correlated with net worth. To save on space, we defer this additional evidence to Appendix B.3.

We attribute the part of the observed increase in equity investments due to diversification in private equity investments in the top panel of Figure 5 to scale dependence in the model. We view this choice as conservative as it constitutes a *lower bound* on the effect of wealth on the equity share of investors. To match this increase within the model, we specify the portfolio share as:

$$\omega(a,\vartheta) = \vartheta\Big(\underline{\omega} + \varpi(a)\Big), \qquad \varpi(a) = \min\Big\{\omega_1\left(\max\left\{a - \underline{a}_{\omega}, 0\right\}\right)^{\gamma_\omega}, \omega_2\Big\}, \tag{31}$$

where $\underline{\omega} = 0.4$ is the average share invested in equity, conditional on being an investor.[36] From Figure 5, $\underline{a}_{\omega}$ is chosen to correspond to the wealth level of the 70^{th} wealth percentile in the model, such that there is no scale dependence in risky portfolio observed below this percentile. The level and the shape parameters are endogenously set to $\omega_1 = 0.072$ and $\gamma_\omega = 0.30$ to replicate the average share invested in risky equity (through additional investments) of 11% for households within the [95-97.5] wealth quantile, and of 20% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.20$ matches the average risky portfolio share above the top 0.1%.

In Appendix C.1, we report the model fit regarding the scale and type dependence parameters.

[35] This result is not due to composition effects. Even when focusing on single households, private equity investments increase with wealth through diversification. Moreover, additional businesses are different than the first established business: 70% of additional private equity investments are made in a different sector. Additionally, they are only slightly more represented in finance-related industries, thereby limiting the concern that it may constitute a financial affiliate company. Those facts challenge the widely held view that business owners are poorly diversified (Moskowitz and Vissing-Jørgensen, 2002) and are consistent with Penciakova (2018). Diversification occurs, but only at the very top. In numbers, 12% of business owners own multiple managed businesses in the US. Among the top 1%, however, this number increases to 40%.

[36] In an alternative calibration strategy, we used longitudinal information of returns to wealth in the PSID to calibrate portfolio shares $\omega(a,\vartheta)$ such that they are consistent with the shape of returns to net worth. However, it is difficult, given the small number of observations in the PSID, to test for non-linearity of scale-dependence at the top of the distribution. Moreover, empirical evidence suggests that scale-dependence occurs in both the extensive and the intensive margins of equity investments, thus introducing important interactions which are captured within our specification.

5 Properties of the Model

Before turning to the main wealth taxation experiment, we first discuss key properties of the calibrated model regarding wealth inequality and returns to wealth. To isolate the driving forces behind the results of our benchmark model, we describe versions of our model with various combinations of type and scale dependence. In these versions, parameters are *always* recalibrated to match the same targets as in the benchmark economy. We find that several combinations of both dependencies deliver close to observationally equivalent cross-sectional moments with respect to wealth and return distributions, but they imply distinct aggregate responses to a wealth tax.

Comparison with alternative models We will subsequently compare our benchmark model (denoted M1) with the following model versions:

(M2) A pure scale dependence model (scale-model) in line with the information acquisition model of Peress (2004) and the incomplete markets models of Meeuwis (2019) and Hubmer et al. (2020). In this version, we shut down type-dependence. All households in the economy now invest in risky assets (i.e. $\vartheta = 1$), yet the amount depends on wealth only. Parameters ω_1, ω_2 and γ_ω match the average portfolio of risky equity observed in the SCF data (cf. Figure 3).

(M3) A pure type dependence model (type-model) that resembles a stylized version of the capitalist/entrepreneur framework along the lines of Moll (2014) or Gomez et al. (2016). There are no scale dependence effects, i.e. the wealth-dependent propensity to select as an investor is set to $\lambda(a) = 0$, the wealth-dependent portfolio component to $\varpi(a) = 0$, and parameters $\overline{\pi}_\vartheta$ and $\underline{\omega}$ are recalibrated to match the fraction of investors and the average portfolio share.

(M4) Despite our focus on type versus scale dependence, one may ask how a version of the widely used capitalist/entrepreneur framework (Cagetti and De Nardi, 2006; Kitao, 2008; Guvenen et al., 2019) with heterogeneity in household investments and returns compares to our benchmark model and the data. The key difference relative to our benchmark stems from the assumption of decreasing returns to scale (DRS) on private equity investments. For this reason, we denote this alternative version the type–DRS entrepreneur model. Apart from this, there is limited scale dependence in equity investment, as those investments are often tight to a borrowing constraint proportional to wealth.[37] To closely replicate such a model version, we impose a risky asset investment share $\underline{\omega} = 1$, no return risk $\sigma_\kappa = 0$, DRS on the entrepreneurial technology $\nu_I = 0.9$, a transition matrix of entrepreneur skill-type ϑ taken

[37] To be precise, those frameworks may feature additional scale dependence in the selection into private equity business investments through an occupational choice (worker versus entrepreneur). A given entrepreneur, however, can invest at most a fixed fraction of her wealth, $k \leq \lambda a$, where λ captures the tightness of the borrowing constraint. In Cagetti and De Nardi (2006), this borrowing limit is endogenous, but turns out to be quasi-linear in wealth.

from Cagetti and De Nardi (2006) to obtain a fraction of entrepreneurs of 8%.[38]

(M5) A no return heterogeneity model that is characterized by $\omega(a, \vartheta) = 0$. In this version, it is not possible to replicate the observed high concentration of wealth at the very top.

In each version, ϕ and β are recalibrated to match the $\frac{K}{Y}$ ratio and the top 1% wealth share.

5.1 Wealth Inequality and Returns to Wealth

As demonstrated in the analytical framework of section 2, it is important that the quantitative model captures well the shape of the wealth distribution, especially at the very top, as this conditions the relative strength of type and scale dependence effects. In Table 6, we first assess the models' accuracy in generating a realistic wealth distribution relative to its empirical counterpart. A striking result is that, beyond the targeted top 1% wealth share, the benchmark model and the alternatives (i.e. models M1 to M4) account remarkably well for the empirical top wealth shares. We do not consider the ability of our model to reproduce the wealth distribution as a success *per se*, since return heterogeneity has been shown to generate high concentration of wealth (Benhabib et al., 2011). However, the fact that our benchmark economy indeed successfully replicates wealth inequality allows us to appropriately study tax experiments that are highly redistributive across the wealth distribution. Moreover, our results point to the observation that type and scale dependence may not be distinguishable based on their ability to generate high inequality, as both mechanisms actually deliver a good fit of top wealth shares.[39]

Table 4. Wealth distribution in the data and models.[a]

		Gini [c]	Share of wealth (in %) held by the top x%						
			40	20	10	5	1	0.1	0.01
US data (World Inequality Database)		0.82	97.5	85.1	70.6	57.7	35.5	18.0	9.0
US data (adjusted SCF)[b]			97.2	86.4	72.7	59.7	37.2	17.8	7.3
M1	benchmark model	0.80	93.4	84.2	71.9	59.3	35.4	18.2	8.9
M2	scale model	0.82	94.6	85.7	73.6	60.3	35.2	20.7	11.7
M3	type model	0.78	92.5	82.0	67.1	56.2	35.7	20.2	10.9
M4	type–DRS entrepreneur model	0.78	93.3	82.0	68.9	56.6	35.9	14.7	4.9

[a] The top one percent wealth share is targeted.
[b] Adjusted for under-representation and underreporting using the procedure in Vermeulen (2016).
[c] The wealth Gini is based on the average estimated from the SCF waves from 1989 to 2019.

This high concentration of wealth can be traced back to substantial heterogeneity in wealth

[38] In our view, this specification is close to the traditional entrepreneur/capitalist type of model where entrepreneurs invest the total amount of their assets in private equity investments subject to a DRS technology. It induces an optimal amount of equity that a household would like to invest. This maximum is never reached under our parameterization. Notice that DRS is often imposed as a relevant assumption in the firm dynamics literature (Lucas Jr, 1978).

[39] This observation goes back to Benhabib, Bisin and Luo (2019) (Table 9). They find that including scale dependence in returns to wealth in a model with type dependence does not provide further explanatory power on the model ability to match top wealth inequality.

returns implied by the different equity portfolio allocations among households. Table 5 shows that consistent with estimates from various data sources, the benchmark model and the alternatives produce average returns to wealth which increase along the wealth distribution.

In the type-model (M3), the increase in returns to wealth is driven by selection only. As investment skill-types are persistent, households with a high propensity to invest in equity have higher expected returns for several periods and are thus more likely to be represented at the top of the distribution, hence driving the observed cross-sectional relationship.[40] In the scale-model (M2), the relationship is generated intuitively, as higher levels of wealth are associated with higher risk-taking and higher expected returns. In the benchmark model (M1), both scale and type dependence drive the observed pattern. To see this, we report the average returns to wealth across the wealth distribution assuming that the *pure orthogonal* type component in (43) is zero, i.e. $\underline{\omega} = 0$, and assuming no scale dependence on the intensive margin, i.e. $\varpi(a) = 0$. In line with the aforementioned empirical evidence in Fagereng et al. (2020) and Bach et al. (2020), we find that both forces shape average returns along the wealth distribution. In contrast, in the type–DRS model (M3) the scale dependence shifts sign due to the DRS specification, and the overall shape of returns across the wealth distribution is now hump-shaped, i.e. it decreases at the very top. Such a negative dependence in returns is not observed in data.[41] Finally, it should be noted that the idiosyncratic *luck* component κ contributes to the overall wealth inequality (Benhabib et al., 2011). In the benchmark model, the standard deviation of returns to wealth is 12.5%, compared to 15% in the PSID. Finally, in Appendix C.3 we show that the 5 years wealth mobility matrices from the models M1 – M3 are comparable to the one obtained from the PSID.

Decomposition In Table 6, we then ask how much type and scale dependence contribute to the observed wealth inequality in the calibrated benchmark model. Specifically, we shut down one component at a time and recompute the stationary distribution without this component, keeping everything else unchanged. Counterfactual (D1) isolates the orthogonal type dependent component by assuming $\underline{\omega} = 0$. Counterfactual (D2) isolates the effects of scale dependence by shutting down both intensive and extensive margin scale effects. Lastly, type and scale dependence in portfolio choices are jointly shut down in counterfactual (D3). Both type and scale dependence are important drivers of wealth inequality, as evidenced by the lower top wealth shares under those counterfactuals. As expected, a model without heterogeneity in capital investments fails to account for the high wealth concentration observed in the data. In such cases, the tail of the wealth distribution inherits the (lower) tail of the labor income distribution. In the model, the income Pareto tail is 1.9 against 1.4 for the empirical wealth distribution from the adjusted SCF.

[40] See Benhabib et al. (2011) and Moll (2014) for a theoretical illumination on the role of persistence in capital returns.

[41] In Bach et al. (2020), there is a slight decrease in private equity returns with respect to wealth due to leverage.

Table 5. Mean returns to wealth (in %) along the wealth distribution: data and model.

Wealth group	Data[a]				Model						
	PSID	SCF	Norway	Sweden	benchmark[b]			scale	type	type–DRS	
					M1	type $\varpi(a)$=0	scale $\underline{\omega}$=0	M2	M3	M4	scale effect
P40-P50	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
P50-P60	−0.6	|	∼ 1.0	0.2	0.0	0.0	0.3	0.0	0.1	0.1	−0.2
P60-P70	−0.9	−0.4		0.3	0.3	0.3	0.6	0.0	0.6	0.7	−0.5
P70-P80	−0.8	0.0	∼ 2.5	0.3	0.9	0.9	1.0	0.3	1.3	1.6	−0.7
P80-P90	0.5	0.2		0.5	1.7	1.7	1.6	1.4	2.1	2.7	−1.0
P90-P95	3.8	1.4	∼ 4.0	0.8	3.6	1.5	2.0	3.8	2.7	3.4	−1.4
P95-P97.5	5.8	2.6	∼ 6.0	1.1	4.6	1.3	3.3	5.9	2.9	5.6	−1.6
P97.5-P99	6.9	3.8		1.5	7.9	3.9	4.0	7.8	7.4	9.9	−1.9
Top 1%	9.6	4.6	∼10.0	2.5	12.5	7.0	5.5	12.2	9.8	7.0	−2.4

Note: "REF" stands for reference wealth bracket, i.e. returns are computed as the difference to the REF.
[a] Estimates are our own for the PSID. They are taken from Xavier (2020) for the SCF, from Bach et al. (2020) for Sweden and from Fagereng et al. (2020) and Halvorsen et al. (2021) for Norway.
[b] The returns are computed assuming $\underline{\omega} = 0$ in the no type model and $\varpi(a) = 0$ in the no scale model.

Table 6. Wealth distribution under alternative model counterfactuals.[a]

		Gini	Share of wealth (in %) held by the top x%						
			40	20	10	5	1	0.1	0.01
	benchmark model	0.80	93.4	84.2	71.9	59.3	35.4	18.2	8.9
D1	no pure type dependence, $\underline{w} = 0$	0.76	93.5	81.3	68.0	48.5	17.0	3.3	0.1
D2	no scale, $\lambda(a) = \varpi(a) = 0$	0.75	91.1	77.8	63.9	51.5	31.7	16.3	8.0
D3	no portfolio heterogeneity	0.63	86.4	68.2	48.4	31.3	8.7	1.1	0.0

5.2 Wealth Inequality – Output Relationship

We now analyze the response to a *permanent* wealth redistribution. Conditional on the strength of type and scale dependence, our goal is to give a sense of how the alternative model versions locate relative to a situation in which inequality is neutral (cf. the GIF in Figure 1). To do so, we compute the long-run effects of a 1 percent tax levied on the wealth of the top 1% wealthiest households. We report the responses in Table 7.

While models M1–M4 produce close to observationally equivalent wealth distributions, they substantially differ in terms of output response. Moving from the scale (M2) to the type (M3) model reduces output losses from -1.19 percent to -0.64 percent. Under scale-dependence, risky asset holdings and thus future wealth are a function of current wealth level itself. This generates a quantitatively strong behavioral dynamic self-enforcing multiplier, which leads to lower returns of the richest households and thus lowers top inequality. In the Type model, the responses are an order of magnitude lower as individuals continue to invest a given share of their wealth into equity, and thus experience high capital returns. The benchmark model (M1), a composite of positive type and scale dependence, falls in between the previous two alternatives. Interestingly, the type–DRS entrepreneur model (M4) generates a lower response, as the negative scale dependence coming from the DRS assumption counterbalances the positive type dependence arising from the sorting of skilled entrepreneurs at the top of the distribution. Finally, a model without portfolio heterogeneity (M5) (a standard Aiyagari (1994) type of model) produces a slight reduction in output of −0.10 percent and is much closer to growth neutrality. In that case, the response comes from a reduction of capital accumulation (K) and is small due to the quasi-linearity of saving decisions for individuals sufficiently far away from the borrowing constraint.

Table 7. Responses to a permanent 1% wealth tax levied on the top 1% wealthiest households.

		ΔGDP (in %)	Δ Top 1% (in pp deviation)	Semi elasticity
M1	benchmark	−0.77	−1.82	0.42
M2	scale	−1.19	−3.50	0.34
M3	type	−0.64	−1.77	0.36
M4	type–DRS	−0.36	−1.48	0.24
M5	no portfolio heterogeneity	−0.10	−0.80	0.13

Additional validation from cross-country evidence In the Online Appendix OA 3.1, we revisit the empirical cross-country relationship between inequality and GDP growth. We extend previous results relying mostly on the Gini coefficient and top income shares to a sample of 29 developed countries by complementing existing estimates of wealth concentration measures with own estimates constructed based on survey data.[42] An increase by 1 percentage point of the top 1% wealth share is associated with a 0.27 percent increase in the subsequent five years average GDP growth.[43] Using data from the most recent Penn World Table, a decomposition of the inequality-growth relationship shows that GDP growth responses to changes in the top 1% wealth shares are mainly reflected in changes of the Solow residual and physical capital accumulation, while the relation is not significant regarding human capital. These findings are in line with the reallocation channel between productive capital investment outlined in this paper.

Using estimates in Table 7, we find that in the type (M3) and scale (M2) dependence models, a 1 percentage point decrease in the top 1% wealth share is respectively associated with a decrease in long-run GDP of 0.36 and 0.34 percent. In contrast, in absence of portfolio heterogeneity (M5), the relation substantially falls, to 0.13 percent. It is worth noting that the benchmark model (M1) produces a slightly stronger association. Among other things, this may be due to the fact that

[42]A wide range of empirical papers studies the link between inequality and growth. For instance, Forbes (2000), Barro (2000) and Halter et al. (2014) use the income Gini coefficient as measure of inequality, while Barro (2008) and Voitchovsky (2005) use quintile and decile income shares. None of the previous papers explore the *Inequality-Growth-slope* using wealth concentration measures. To the best of our knowledge, only Voitchovsky (2005) and Frank (2009) look at the impact of income concentration at different quantiles.

[43]Note that we do not claim any causal relationship in here. Even in the model, a negative wealth shock at the top of the distribution affects inequality and GDP, and their combined change in turn feeds back into inequality and GDP. Therefore, it is hard to identify causality even based on our simulated results.

return heterogeneity may imperfectly reflect productivity differences due to some forms of rent-extraction, i.e. $\mu < 1$. In the next section, we study optimal wealth taxation while we allow returns to imperfectly reflect investment productivity.

Implications for the dynamics of wealth inequality While Gabaix et al. (2016) show that type and scale dependence account for the recent rise of income/wealth inequality, their analysis is uninformative on whether the two are equivalent when both channels are calibrated to match the same empirical moments. Our results substantiate the idea that various degrees of type and scale dependence that are consistent with portfolio and return heterogeneity produce large difference when it comes to understanding the dynamics of wealth inequality. We find that the response of the top 1% wealth share to a wealth tax in the scale model (M2) is twice as large as the one found under the type model (M3). This may have large consequences, as Hubmer et al. (2020) find that changes in top capital income taxes are a key driver of the recent rise in wealth inequality observed in the US. In their words, *"the marked decrease in tax progressivity is by far the most powerful force for the cumulative increase in wealth inequality"*. However, this result is derived based on a model that features scale dependence in portfolio choices only. In the Online Appendix OA 2.1, we support our results in table 7 and show that changes in capital income taxes in the US since 1980 have substantially large differences on the dynamics of wealth inequality depending on whether returns are driven by scale or type dependence.

6 Wealth Taxation: the Role of Type and Scale Dependence

We now proceed to our main experiment and assess the quantitative implications of taxing household wealth at the steady state. We conduct our experiment in three steps. First, we compute the optimal wealth tax in our benchmark economy and decompose the resulting welfare gains. We find that a positive wealth tax above the top 80^{th} percentile is optimal. Second, we deviate from the benchmark calibration and study cases for which returns to wealth imperfectly reflect differences in capital productivity. We numerically argue that the existence of rents does not substantially change the welfare-maximizing tax rate relative to the one obtained under the benchmark calibration. Third, we dissect the key underlying forces behind our results and unravel the distinct effects arising from type and scale dependence.

Thought experiment We assume that tax rates cannot be personalized. The government uses a restricted class of wealth tax functions described by

$$t_a(a;\tau_a,\underline{a}_{max}) = \mathbb{1}_{a\geq\underline{a}_{max}}\tau_a(a-\underline{a}_{max}), \tag{32}$$

and optimizes over the exemption level $\underline{a}_{max}$ and the marginal wealth tax rate τ_a. Imposing restrictions on the class of wealth tax functions the government can choose from is necessary to ensure that the maximization objective is computationally feasible.[44] In equilibrium, the labor income tax τ_w adjusts to ensure that the government budget is balanced. In the Online Appendix OA 1.6, we provide further results when using alternative tax instruments to balance the government budget.

The main trade-offs shaping optimal redistribution are those identified in section 2. The wealth tax balances: (i) equity by reallocating wealth from households with low to high marginal utilities of consumption, (ii) efficiency as wealth-rich households trigger general equilibrium effects through asset reallocation and thus affect productivity and wages, and (iii) rent-extraction, as whenever $\mu < 1$, higher risky asset investments lead to an overall downward adjustment in the aggregate component $\underline{r}$ of returns to wealth.

The criterion to rank different wealth tax functions is based on a consumption-equivalent variation (henceforth CEV) approach. After solving for the stationary equilibrium of a specific tax reform $(\tau_a, \underline{a}_{max})$, we compute the variation Δ^{CEV} of consumption that makes every household in the post-reform economy on average as well-off as in the pre-reform economy. Under this *utilitarian* criterion, aggregate welfare in the *post*-reform economy, $\mathcal{W}^{post}(\tau_a, \underline{a}_{max})$, has to be equal to aggregate welfare in the status quo economy without wealth tax $\mathcal{W}^{pre}(\Delta^{CEV})$, where optimal consumption has been changed by Δ^{CEV} percent, such that

$$\mathcal{W}^{post}(\tau_a, \underline{a}_{max}) = \mathcal{W}^{pre}(\Delta^{CEV}),$$

$$\int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \widetilde{\beta}^t u\Big(c_t^{post}(\mathbf{s}), \ell_t^{post}(\mathbf{s})\Big) \right] d\mathcal{G}^{post}(\mathbf{s}) = \int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \widetilde{\beta}^t u\Big((1+\Delta^{CEV}) c_t^{pre}(\mathbf{s}), \ell_t^{pre}(\mathbf{s})\Big) \right] d\mathcal{G}^{pre}(\mathbf{s}),$$

where $\widetilde{\beta} = \beta(1 - d_j)$. Given our time-separable utility function, it is straightforward to show that the government problem can be stated as follows

$$\arg \max_{\{\tau_a, \underline{a}_{max}\}} \Delta^{CEV}(\tau_a, \underline{a}_{max}) = \left[\frac{\mathcal{W}^{post}(\tau_a, \underline{a}_{max}) - \mathcal{W}^{pre}(0)}{\int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t (1-d_j)^t \frac{c_t^{pre}(\mathbf{s})^{1-\sigma}}{1-\sigma} \right] d\mathcal{G}^{pre}(\mathbf{s})} + 1 \right]^{\frac{1}{1-\sigma}} - 1 .$$

This welfare criterion is widely used in the quantitative macroeconomics literature (among many others Conesa et al. (2009), Guvenen et al. (2019) or Brüggemann (2021)). In our setting, this criterion relies, however, on a "representative utility" that ranks people according to the same preferences whatever the underlying scale or type dependent mechanism is. As such, we do not claim that this criterion captures all relevant welfare effects associated with a wealth tax, which may arise in reality due to a particular scale or type dependent mechanism. Our criterion implic-

[44] Furthermore, many developed countries adopted this wealth tax schedule. A typical wealth tax features an exemption level ranging from the wealth level corresponding to the top 50% (in Switzerland) to the top 5 to 1% wealthiest households (in France).

itly assumes, for instance, that type and scale dependence do not alter household preferences. It should thus be understood as a transparent way to compare welfare consequences among various economies with different degrees of type and scale dependence without actually changing the underlying planner objective.

6.1 Results

The optimal wealth tax system in the benchmark economy is given by a positive marginal tax rate of 0.82 percent with an exemption level of \$550K. To put these numbers into context, the reform is equivalent to imposing a wealth tax on the top 20 percent wealthiest households, that currently hold approximately 85 percent of total wealth in the US economy.

In Table 8, we report that under the optimal tax reform capital and productivity fall below the level of the benchmark economy. Consequently, aggregate output falls as well. This is an immediate implication of the wealth tax, which disproportionately concerns individuals contributing to risky and more productive assets. Additionally, their saving rate drops substantially. The tax reform also induces adjustments in aggregate labor supply arising from two opposite forces. First, wealth-rich households become richer over time which induces a negative wealth effect on labor supply. Second, wealth-rich households become poorer, which induces them to work more. Finally, notice that under the wealth tax, the top 0.1% wealth share increases.

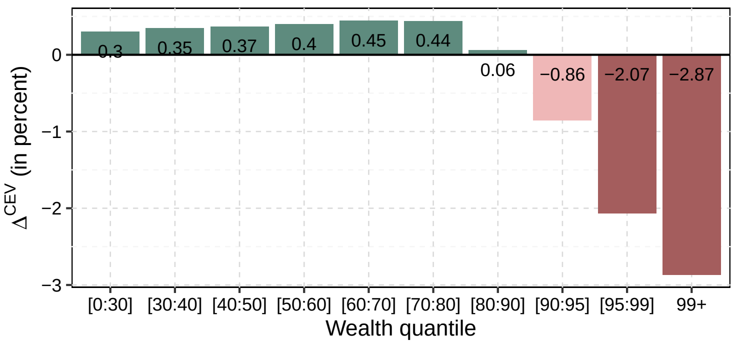
Table 8. Aggregate variables after optimal tax reform.

	Percent change relative to status quo
Total labor supply L	0.26
Capital stock X	-6.19
Productivity X/K	-0.08
Output Y	-2.10
Top 0.1% wealth share (in pp)	0.50
Δ^{CEV}	0.14

6.1.1 Decomposing the welfare gains

The consumption equivalent variation in response to the optimal tax reform is modest with an average of 0.14 percent of yearly consumption. This number, however, conceals important heterogeneous effects across the wealth distribution. In Figure 6, we document that the largest drop in terms of CEV occurs within the top 1% wealth bracket, in which wealth is largely above the exemption level. In contrast, welfare gains of wealth brackets below the 70^{th} wealth percentile are on average positive, in between 0.3 to 0.45 percent of yearly consumption. Furthermore, the CEV is slightly hump-shaped across the wealth redistribution, as revenues that are raised from the wealth tax are balanced with a lower labor tax which disproportionately benefits households with relatively high labor incomes. Finally, the reform is also politically feasible as the majority of households benefits in terms of CEV.

Figure 6. Consumption equivalent variation Δ^{CEV} across wealth distribution.



6.1.2 The effects of rent-extraction

To what extent do differential returns reflect the productivity of capital investments? So far, our baseline results abstract from rents in private returns to wealth. However, private returns may reflect both the productivity of capital investments (MPK) and some form of rents. For example, Smith et al. (2019a) observe for private businesses that the share of value added allocated to owners has increased over time, irrespective of productivity gains. Incorporating this feature into a quantitative exercise is challenging as it requires detailed data allowing to link household investments to their marginal productivity, which must then be compared to private returns of the corresponding household.

Despite this limitation, we show how the presence of a realistic amount of rent extraction would alter the top marginal wealth tax rate. Specifically, we distinguish two types of risky assets. On the one hand, rent-seeking investments, and on the other hand, investments linked to higher productivity and output.[45] Lockwood et al. (2017) report that an increase in the aggregate income share of finance service and law sectors is associated with a decrease in aggregate income. They interpret this finding as indirect evidence for a negative externality from those sectors on aggregate income. Using their estimates, Rothschild and Scheuer (2016) derive an optimal labor income tax taking into account rent-seeking. Following their lead and for the sake of illustration, we make the assumption that returns to wealth obtained from investments in finance and law sectors are associated to rent-seeking activities.

To calibrate the degree of returns associated to rent extraction motives, we use our SCF sample and compute the average share of household equity investments into law and finance sectors between 1998 and 2019. Those equity investments account for roughly 20% of total equity, which

[45]It should be noticed that both Piketty et al. (2014), Rothschild and Scheuer (2016) and Lockwood et al. (2017) focus mainly on rent-seeking from labor income. Moreover, the recent studies by Piketty et al. (2014) and Lockwood et al. (2017) focus on the case where externalities from rent-seeking reduce everyone else's income in a lump-sum fashion rather than the proportional reduction that we consider here through $\underline{r}$. Relatedly, Scheuer and Slemrod (2021) discuss the role of rent-extraction in capital returns.

justifies our choice to set the return wedge to $\mu = 0.8$.[46] Again, the aggregate return component $\underline{r}$ adjusts in equilibrium to ensure that total revenues obtained by households coincide with total revenues distributed by the intermediate good producer. As a higher return wedge, i.e. a lower value of μ, increases the dispersion of returns between households, we recalibrate the values for ϕ and β such that the model matches the top 1% wealth share and a capital-output ratio of 2.6.

To have a direct comparison to our baseline results without rent extraction, i.e. the case of $\mu = 1$, we preserve the wealth exemption threshold to the wealth level corresponding to the 80^{th} wealth percentile, hereafter referred to as $\underline{a}_{max} = F_a^{-1}(0.80)$, and optimize our welfare criterion over the marginal tax rate τ_a. We find that the optimal top marginal wealth tax rate increases slightly relative to the case without rent-extraction, to a rate of 0.92 percent. Implementing this tax reform leads to overall welfare gains equivalent to 0.2 percent of yearly consumption. Therefore, the marginal wealth tax rate *slightly increases* in the degree of rent extraction. In fact, simulating various model versions with different degrees for $\mu < 1$ shows that the marginal tax rate τ_a monotonously increases in the return wedge. This implies that our benchmark tax rate of 0.82 percent, derived when private returns from investment coincide with their associated marginal productivity, constitutes a conservative *lower bound*.

6.2 Dissecting the Effects from Type and Scale Dependence

We now isolate the driving forces behind our two main quantitative results while fixing the wealth exemption level, that is

(A) the optimal marginal wealth tax rate is positive,

(B) the optimal marginal wealth tax rate *slightly increases* in the degree of rent-extraction.

Subsequently, we demonstrate that results *(A)* and *(B)* depend on the quantitative importance of type and scale dependence, and how both mechanisms interact with the extent to which returns to wealth reflect the productivity of capital investments. To unravel the driving forces, we compare the aggregate equilibrium statistics relative to the status quo along several model alternatives where various combinations of type and scale dependence and the return wedge μ are adopted. Under all alternatives, the parameters ϕ and β are *always* recalibrated to match the same targets regarding the top 1% wealth share and the capital-output ratio.

[46]The share of equity invested into both sectors displays substantial heterogeneity across the wealth distribution. Notably, it is particularly important at the top of the wealth distribution, from 15% at the bottom 99% to 22% for the top 1% wealthiest households. In an alternative experiment, we capture this non-linearity by assuming that the return wedge is wealth dependent, i.e $\mu(a) = \mu_1 a^{\mu_2}$, for some positive parameters μ_1 and μ_2. We find similar results.

6.2.1 The role of type and scale dependence

We first shed light on the role of type and scale dependence in driving our quantitative results. We compute the aggregate implications of taxing wealth at the optimum of the benchmark economy, characterized by $\mu = 1$, $\tau_a = 0.0082$ and $\underline{a}_{max} = F_a^{-1}(0.80)$, in the scale model (M2) and in the type model (M3). Table 9 displays the aggregate responses. A striking result is that scale and type models exhibit opposite responses to a positive wealth tax with respect to aggregate labor supply and productivity. Moreover, the welfare gains turn out to be significant and positive under the type model, and negative under the scale model.

Table 9. Aggregate variables after the optimal tax reform derived under the benchmark economy.

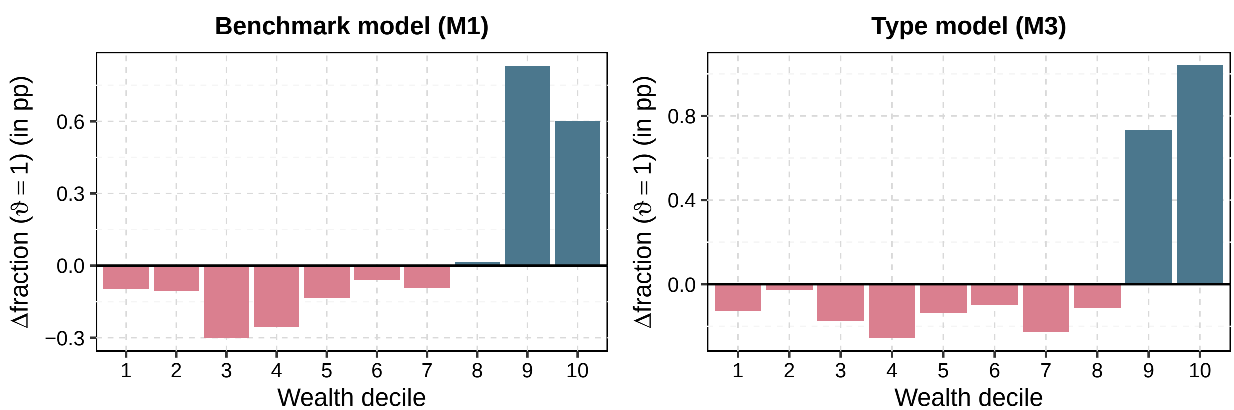
	Percent changes relative to status quo		
	scale model (M2)	benchmark model (M1)	type model (M3)
Total labor supply L	0.50	0.26	−0.08
Capital stock X	−7.33	−6.19	−4.84
Productivity X/K	−1.55	−0.08	1.27
Output Y	−2.39	−2.10	−1.82
Top 0.1% wealth share (in pp)	−0.15	0.50	0.16
Δ^{CEV}	−0.36	0.14	0.52

Two forces rationalize the above findings. First, the scale model (M2) triggers important general equilibrium effects in response to wealth taxation. As shown in Table 9, this manifests in relatively large output losses relative to the type model (M3). The reason is that risky investment behavior is itself a function of wealth, and thus any change in wealth is accompanied by strong contemporaneous behavioral responses in terms of portfolio allocation and savings which transmit across periods. This *snowball* effect induced by the wealth tax reduces wealth accumulation and the amount of risky asset investments undertaken by wealth-rich households, translating into permanent lower productivity and equilibrium wage w. Nevertheless, households work more in the post reform equilibrium, as they face lower marginal labor income taxes and the income effect is sufficiently strong compared to the substitution effect. Finally, these efficiency losses generated by the aforementioned general equilibrium effects outweigh the equity gains from redistribution such that the average consumption equivalent variation turns out to be negative.

Second, under the type model (M3), we find opposite effects on labor, productivity and welfare gains. To understand the underlying mechanisms, it is important to bear in mind that the persistence of types is crucial for engendering a selection of households with high returns more frequently into the top of the wealth distribution, as they experience high returns to their wealth during several consecutive periods. When wealth-rich households face a tax on the stock of their wealth, those with high returns to wealth are relatively less affected by the wealth tax. As a result, they dissave at a lower rate relative to wealth-rich households who invest in less risky assets and experience lower returns to wealth.[47] Therefore, by taxing the stock of wealth of the richest households, the government creates an environment where *only the fittest survive at the top*. Wealth taxation thus reinforces the selection of agents with higher capital income – associated with higher capital productivity – even further. In Figure 7, we show that the wealth tax indeed selects a higher fraction of investors at the top, both in the benchmark model (M1) (left panel) and the type model (M3) (right panel). Note that the hump-shaped pattern at the top in the benchmark economy comes from the scale dependence effects on the risky investment participation margin. In the end, highly skilled investors hold a higher fraction of total wealth, which raises productivity. This effect outweighs the negative effects of a lower capital accumulation on GDP such that welfare gains are large relative to the ones obtained in the benchmark economy.

Due to those two opposing forces from scale and type dependence, aggregate productivity is approximately irresponsive to the implementation of a wealth tax in the benchmark economy. In the next section, we show that these distinct implications on the aggregates also shape the optimal marginal wealth tax rate.

Figure 7. Fraction of high investor type ($\vartheta = 1$) relative to the status quo per wealth decile.



6.2.2 Optimal wealth tax under type and scale dependence

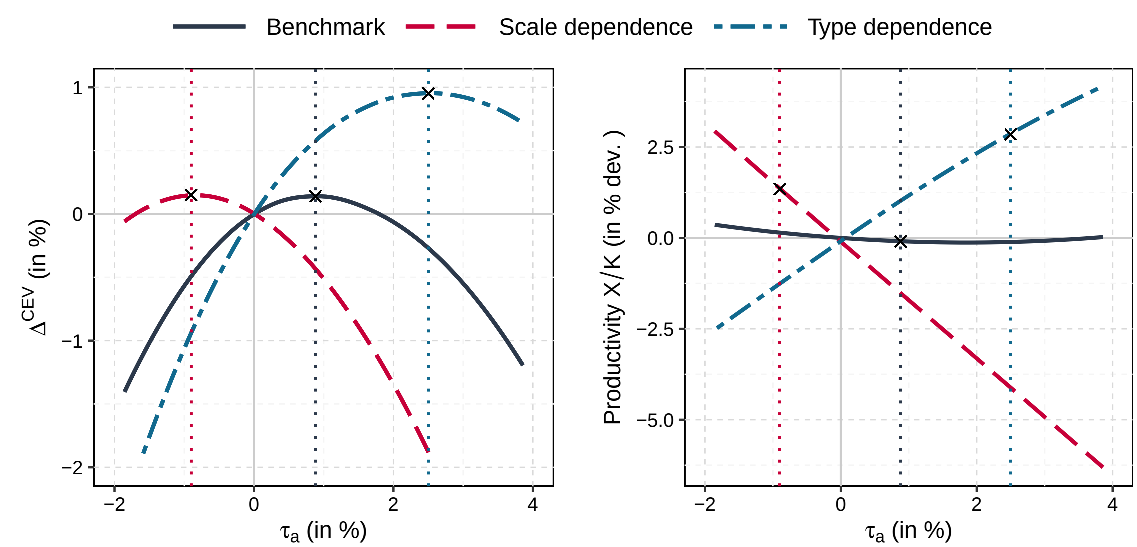
In our second experiment we optimize over the marginal wealth tax rate under the scale model (M2) and under the type model (M3), keeping the wealth exemption level constant such that it corresponds to the 80^{th} percentile. As shown in the left panel of Figure 9, the optimal marginal wealth tax rate under the scale model is negative, at a rate of -0.88 percent, while it is substantially positive under the type model, at a rate of 2.41 percent. Which mechanisms explain these different implications for optimal redistribution under the two polar cases?

[47] The fact that higher returns lead individuals to save at a higher rate is a reminiscence of previous findings in the literature, in which entrepreneurs with high returns on their capital save at a higher rate relative to workers with low returns (see, for instance, Cagetti and De Nardi (2006)). This point is also formalized in Fagereng et al. (2019) in the absence of return risk.

In the right panel of Figure 8, we plot the capital productivity, i.e. the ratio X/K, as a function of the marginal wealth tax rate τ_a. Under type dependence, the change in the selection of skilled types along the wealth distribution leads to an increase in productivity despite the negative effects on capital accumulation. The welfare maximizing wealth tax rate is large with sizable welfare gains equivalent to 0.96 percent of yearly consumption. The reasoning behind this finding is in line with the optimal wealth tax result stated in Guvenen et al. (2019), in which return heterogeneity stems mostly from heterogeneous entrepreneurial skills among households, i.e. from type-dependence. In contrast, scale dependence triggers a strong behavioral response on risky asset investments, such that productivity decreases in the marginal wealth tax, and so do welfare gains as well.

Those distinct forces on productivity rationalize Result *(A)* under our benchmark economy: the welfare-maximizing wealth tax rate is positive. In this economy, type and scale dependence outweigh each other and the productivity becomes roughly irresponsive to the implementation of a wealth tax. As such, the optimal welfare-maximizing wealth tax rate is positive but far below the one implied under pure type dependence.

Figure 8. CEV welfare as function of τ_a under type and wealth dependence.



Remark: results are derived by comparing the welfare measure within different long-run stationary economy in which the marginal tax rate τ_a varies.

6.3 The Interaction between Type and Scale Dependence and Rents

We now show how the share of rents in private returns affects the welfare-maximizing wealth tax rate under the scale, the type and the benchmark models (M1–M3). Again, we set the wealth exemption level to the one implied by the optimal exemption level under the benchmark model assuming $\mu = 1$. We then compute the optimal marginal tax rate τ_a in the presence of rent-extraction by fixing $\mu = 0.8$ and recalibrate ϕ and β such that the alternative model versions match a top 1% wealth share of 0.36 and a capital-output ratio of 2.6.

In Figure 9 we show that the optimal wealth tax rate is *increasing* in the presence of pure rents in returns to wealth under the scale model (M2). In contrast, the optimal wealth tax rate is *decreasing* in the presence of pure rents in returns to wealth under the type model (M3). This finding mirrors our productivity result, but goes in the opposite direction concerning welfare.

In the Scale model M2, wealth-rich households invest in risky assets, but only a share of those assets are associated with higher productivity, while the remainder is associated with higher rent-extraction. As discussed earlier, in that case, it is optimal to lower the subsidy relative to the case without rent-extraction. The optimal tax rate τ_a thus rises from -0.88 percent to -0.36 percent.

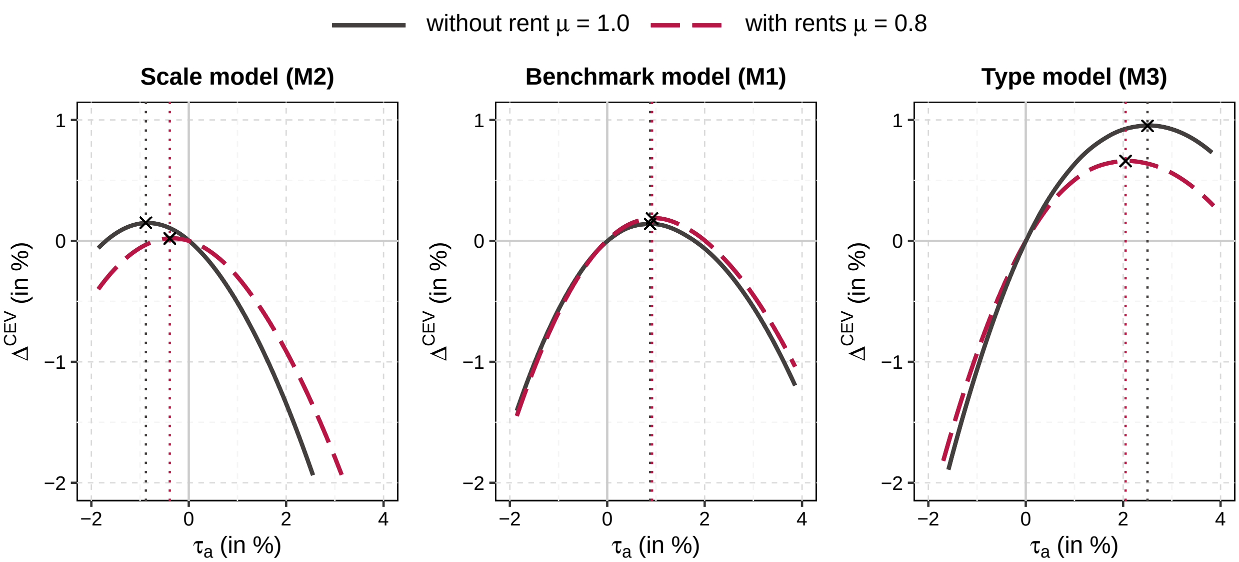
In the type model (M3), highly skilled investors obtain high returns because they systematically invest a higher share of their wealth in riskier investments which now partly reflects productivity *and* rents. As before, taxing wealth leads high return investors to be even more represented at the top, raising in this case both productivity and the size of rents in the economy. Productivity gains are, however, lower than in the baseline case with $\mu = 1$. Therefore, everything else equal, a wealth tax in this economy is less powerful relative to the case of a type model (M3) without rents. On top of this, due to rent-extraction, the stronger selection of high return investors induces a general equilibrium adjustment of the aggregate return $\underline{r}$ for all individuals in the economy, which pushes towards a lower wealth tax. Those two forces lead the optimal wealth tax rate to *decrease* with the size of rents, from a rate of 2.41 percent without rent-extraction to a rate of 2.06 percent when $\mu = 0.8$.

Quite surprisingly, we find that, again, type and scale dependence effects outweigh each other in the benchmark economy, such that the marginal wealth tax rate is almost not responding to the presence of rent-extraction. This leads to Result *(B)*: the welfare-maximizing wealth tax rate is slightly *increasing* in the share of rents in returns.

Finally, notice that our results under pure scale or type dependence with $\mu = 1$, i.e. in the absence of a return wedge, can be viewed as lower and upper bounds on the optimal marginal wealth tax. Our results point out that depending on the relative strength of scale and type dependence consistent with the observed return heterogeneity in the data, the welfare-maximizing tax rate lies in between $[-0.8, 2.4]$ percent.

To summarize, a key take away is that in order to understand the consequences of wealth taxation in an economy where returns to wealth may or may not coincide with capital productivity, it is essential to take into account the relative importance of type and scale dependence in household investments and associated returns to wealth.

Figure 9. CEV welfare as function of τ_a under type and *wealth* dependence: sensitivity rent-extraction.



Remark: results are derived by comparing the welfare measure within different long-run stationary economies in which the marginal tax rate τ_a varies.

6.4 Further Robustness

In addition to the previous analysis, we have also explored the role of particular elements of our quantitative model including (i) the life-cycle structure with the mortality rate $d_j > 0$, the age-dependent earnings component $\zeta_j \neq 1$ and the social security pension ζ_J , (ii) the endogenous labor choice, (iii) the tax instrument used to balance the government budget, (iv) the presence of idiosyncratic return risk. We do so by reassessing the optimal wealth tax in versions of our model where various combinations of these elements are shut down. We found that (i) and (ii) have little influence on our main qualitative message regarding the distinction between type and scale dependence. Concerning (iii), we find that redistribution through a lump-sum tax provides similar results, while redistribution through capital income tax reinforces even further the selection of high types at the top. Removing the risky component κ in returns lowers wealth inequality in the stationary equilibrium. This is compensated by increasing the excess wealth return ϕ to match the top 1% wealth share. More importantly, under all previously discussed model alternatives, we find that our results do not hinge on particular elements. A high positive wealth tax is optimal under type dependence and a low or negative wealth tax is optimal under scale dependence.

7 Conclusion

In this paper, we first develop a conceptual framework to study the macroeconomic and welfare implications of wealth redistribution, unraveling and clarifying the key economic forces behind many heterogeneous agents incomplete markets models. Despite its stylized nature, our analytical two-period model identifies four statistics that are crucial for understanding these implications: (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking to wealth (scale dependence), (iii) the sorting of types along the wealth distribution (type dependence), and (iv) the extent to which returns to wealth reflects investment productivity.

In a second step, we construct a full-blown quantitative model and analyze the key elements that we identified as particularly relevant within our theoretical framework. The model accounts for the highly concentrated wealth and returns distributions through type and scale dependence in portfolio choices. Our model is consistent with empirical evidence from the SCF and PSID. We show that the underlying force behind wealth accumulation and inequality, i.e. type or scale dependence, shapes distinct aggregate responses to a top wealth tax, both qualitatively and quantitatively. The aggregate responses of productivity and inequality are large under pure scale dependence. Under type dependence, however, the joint distribution of investor-types and wealth non-trivially reacts to the implementation of a wealth tax, as a top marginal wealth tax selects high capital income households more effectively into the top of the wealth distribution.

The welfare implications of a wealth tax depend on the degree of scale and type dependence together with the extent to which returns to wealth reflect the productivity of investments. In our benchmark economy, the optimal top marginal wealth tax rate is 0.8 percent above an exemption level of $550K, as long as the model features scale and type dependence consistent with data.

Future research is needed to understand the relationship between the joint distribution of wealth, returns, and portfolio allocation. Specifically, empirical studies are helpful in disentangling scale and type dependence in the decision of agents, and, in turn, to determine the optimal taxation of wealth-rich households. Moreover, future research should attempt to empirically evaluate the pass-through between the productivity of investments and returns to wealth along the lines of Lockwood et al. (2017) or Smith et al. (2019a). Finally, we abstracted from tax avoidance motives. Substantial amounts of wealth are held abroad (Alstadsæter et al., 2018), and a wealth tax may foster incentives for tax avoidance. Such considerations induce a form of negative scale dependence that our quantitative model may additionally consider. These are important and, to a large extent, unexplored issues that we leave for future work.

References

AIYAGARI, S. RAO (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, Vol. 109, pp. 659–684.

AIYAGARI, S RAO (1995): "Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting," *Journal of Political Economy*, Vol. 103, pp. 1158–1175.

ALPANDA, SAMI AND GEOFFREY WOGLOM (2007): "The Case Against Power Utility and a Suggested Alternative: Resurrecting Exponential Utility," MPRA Working Paper No. 5897.

ALSTADSÆTER, ANNETTE, NIELS JOHANNESEN, AND GABRIEL ZUCMAN (2018): "Who owns the wealth in tax havens? Macro evidence and implications for global inequality," *Journal of Public Economics*, Vol. 162, pp. 89–100.

ANGELETOS, GEORGE-MARIOS (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, Vol. 10, pp. 1–30.

ANGELETOS, GEORGE-MARIOS AND LAURENT-EMMANUEL CALVET (2006): "Idiosyncratic Production Risk, Growth and the Business Cycle," *Journal of Monetary Economics*, Vol. 53, pp. 1095–1115.

BACH, LAURENT, LAURENT E CALVET, AND PAOLO SODINI (2020): "Rich pickings? Risk, return, and skill in household wealth," *American Economic Review*, Vol. 110, pp. 2703–47.

BANERJEE, ABHIJIT V AND ANDREW F NEWMAN (1993): "Occupational choice and the process of development," *Journal of Political Economy*, Vol. 101, pp. 274–298.

BARRO, ROBERT J (2000): "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, Vol. 5, pp. 5–32.

——— (2008): "Inequality and growth revisited,"Technical report, ADB Working paper series on regional economic integration.

BENHABIB, JESS, ALBERTO BISIN, AND MI LUO (2019): "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, Vol. 109, pp. 1623–47.

BENHABIB, JESS, ALBERTO BISIN, AND SHENGHAO ZHU (2011): "The distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, Vol. 79, pp. 123–157.

BEWLEY, TRUMAN (1986): "Stationary monetary equilibrium with a continuum of independently fluctuating consumers," *Contributions to mathematical economics in honor of Gérard Debreu*, Vol. 79.

BOAR, CORINA AND VIRGILIU MIDRIGAN (2019): "Markups and inequality,"Technical report, National Bureau of Economic Research.

——— (2020): "Efficient Redistribution,"Technical report, National Bureau of Economic Research.

BOURGUIGNON, FRANÃ§OIS (1981): "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function," *Econometrica*, Vol. 49, pp. 1469–1475.

BRÜGGEMANN, BETTINA (2020): "Higher taxes at the Top: The role of entrepreneurs," *American Economic Journal: Macroeconomics, Forthcoming*.

——— (2021): "Higher taxes at the top: The role of entrepreneurs," *American Economic Journal: Macroeconomics*, Vol. 13, pp. 1–36.

BRUNNERMEIER, MARKUS K. AND STEFAN NAGEL (2008): "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation," *The American*

Economic Review, Vol. 98, pp. 713–736.

CAGETTI, MARCO AND MARIACRISTINA DE NARDI (2006): “Entrepreneurship, frictions, and wealth,” *Journal of Political Economy*, Vol. 114, pp. 835–870.

——— (2009): “Estate taxation, entrepreneurship, and wealth,” *American Economic Review*, Vol. 99, pp. 85–111.

CALVET, LAURENT, JOHN CAMPBELL, AND PAOLO SODINI (2009): “Fight or Flight? Portfolio Rebalancing by Individual Investors,” *The Quarterly Journal of Economics*, Vol. 124, pp. 301–348.

CALVET, LAURENT E, JOHN Y CAMPBELL, FRANCISCO J GOMES, AND PAOLO SODINI (2019): “The cross-section of household preferences,”Technical report, Working Paper, Harvard University.

CARROLL, CHRISTOPHER D (2006): “The method of endogenous gridpoints for solving dynamic stochastic optimization problems,” *Economics letters*, Vol. 91, pp. 312–320.

CHETTY, RAJ, NATHANIEL HENDREN, PATRICK KLINE, AND EMMANUEL SAEZ (2014): “Where is the land of opportunity? The geography of intergenerational mobility in the United States,” *The Quarterly Journal of Economics*, Vol. 129, pp. 1553–1623.

CONESA, JUAN CARLOS, SAGIRI KITAO, AND DIRK KRUEGER (2009): “Taxing capital? Not a bad idea after all!,” *American Economic Review*, Vol. 99, pp. 25–48.

DE NARDI, MARIACRISTINA AND FANG YANG (2016): “Wealth inequality, family background, and estate taxation,” *Journal of Monetary Economics*, Vol. 77, pp. 130–145.

DIAMOND, ROBERT M (1998): *Designing and Assessing Courses and Curricula: A Practical Guide. The Jossey-Bass Higher and Adult Education Series.*: ERIC.

FAGERENG, ANDREAS, CHARLES GOTTLIEB, AND LUIGI GUISO (2017): “Asset market participation and portfolio choice over the life-cycle,” *The Journal of Finance*, Vol. 72, pp. 705–750.

FAGERENG, ANDREAS, LUIGI GUISO, DAVIDE MALACRINO, AND LUIGI PISTAFERRI (2020): “Heterogeneity and persistence in returns to wealth,” *Econometrica*, Vol. 88, pp. 115–170.

FAGERENG, ANDREAS, MARTIN BLOMHOFF HOLM, BENJAMIN MOLL, AND GISLE NATVIK (2019): “Saving behavior across the wealth distribution: The importance of capital gains,”Technical report, National Bureau of Economic Research.

FELDSTEIN, MARTIN S (1969): “The effects of taxation on risk taking,” *Journal of Political Economy*, Vol. 77, pp. 755–764.

FLAVIN, MARJORIE AND TAKASHI YAMASHITA (2002): “Owner-occupied housing and the composition of the household portfolio,” *American Economic Review*, Vol. 92, pp. 345–362.

FORBES, KRISTIN (2000): “A Reassessment of the Relationship between Inequality and Growth,” *The American Economic Review*, Vol. 90, pp. 869–887.

FRANK, MARK W (2009): “Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures,” *Economic Inquiry*, Vol. 47, pp. 55–68.

GABAIX, XAVIER (2011): “The granular origins of aggregate fluctuations,” *Econometrica*, Vol. 79, pp. 733–772.

GABAIX, XAVIER, JEAN-MICHEL LASRY, PIERRE-LOUIS LIONS, AND BENJAMIN MOLL (2016): "The dynamics of inequality," *Econometrica*, Vol. 84, pp. 2071–2111.

GALOR, ODED AND JOSEPH ZEIRA (1993): "Income Distribution and Macroeconomics," *Review of Economic Studies*, Vol. 60, pp. 35–52.

GERRITSEN, AART, BAS JACOBS, ALEXANDRA VICTORIA RUSU, AND KEVIN SPIRITUS (2020): "Optimal taxation of capital income with heterogeneous rates of return."

GOLLIER, CHRISTIAN (2001): "Wealth Inequality and Asset Pricing," *The Review of Economic Studies*, Vol. 68, pp. 181–203.

GOMEZ, MATTHIEU ET AL. (2016): "Asset prices and wealth inequality," *Unpublished paper: Princeton*.

——— (2018): "Decomposing the Rise in Top Wealth Inequality," in *2018 Meeting Papers*, No. 1116, Society for Economic Dynamics.

GUISO, LUIGI AND MONICA PAIELLA (2000): "Risk Aversion, Wealth and Financial Market Imperfections," Mimeo.

GUVENEN, FATIH, GUEORGUI KAMBOUROV, BURHANETTIN KURUSCU, SERGIO OCAMPO-DIAZ, AND DAPHNE CHEN (2019): "Use it or lose it: Efficiency gains from wealth taxation,"Technical report, National Bureau of Economic Research.

GUVENEN, FATIH, FATIH KARAHAN, SERDAR OZKAN, AND JAE SONG (2021): "What Do Data on Millions of US Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica*, Vol. 89, pp. 2303–2339.

HALTER, DANIEL, MANUEL OECHSLIN, AND JOSEF ZWEIMÜLLER (2014): "Inequality and growth: the neglected time dimension," *Journal of Economic Growth*, Vol. 19, pp. 81–104.

HALVORSEN, ELIN, JOACHIM HUBMER, SERDAR OZKAN, AND SERGIO SALGADO (2021): "Why are the Wealthiest So Wealthy?".

HEATHCOTE, JONATHAN, KJETIL STORESLETTEN, AND GIOVANNI L VIOLANTE (2010): "The macroeconomic implications of rising wage inequality in the United States," *Journal of Political Economy*, Vol. 118, pp. 681–722.

HEATHCOTE, JONATHAN, KJETIL STORESLETTEN, AND GIOVANNI L. VIOLANTE (2017): "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, Vol. 132, pp. 1693–1754.

HERRANZ, NEUS, STEFAN KRASA, AND ANNE P VILLAMIL (2015): "Entrepreneurs, risk aversion, and dynamic firms," *Journal of Political Economy*, Vol. 123, pp. 1133–1176.

HUBMER, JOACHIM, PER KRUSELL, AND ANTHONY A SMITH JR (2020): "Sources of US wealth inequality: Past, present, and future," in *NBER Macroeconomics Annual 2020, volume 35*: University of Chicago Press.

HUGGETT, MARK (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, Vol. 17, pp. 953–969.

HURST, ERIK AND ANNAMARIA LUSARDI (2004): "Liquidity constraints, household wealth, and entrepreneurship," *Journal of Political Economy*, Vol. 112, pp. 319–347.

İMROHOROĞLU, SELAHATTIN (1998): "A quantitative analysis of capital income taxation," *International Economic Review*, pp. 307–328.

JAKOBSEN, KATRINE, KRISTIAN JAKOBSEN, HENRIK KLEVEN, AND GABRIEL ZUCMAN (2020): "Wealth taxation and wealth accumulation: Theory and evidence from Denmark," *The Quarterly Journal of Economics*, Vol. 135, pp. 329–388.

KACPERCZYK, MARCIN, JAROMIR NOSAL, AND LUMINITA STEVENS (2019): "Investor sophistication and capital income inequality," *Journal of Monetary Economics*, Vol. 107, pp. 18–31.

KALDOR, NICHOLAS (1956): "Alternative Theories of Distribution," *The Review of Economic Studies*, Vol. 23, pp. 83–100.

KAPLAN, GREG, BENJAMIN MOLL, AND GIOVANNI L VIOLANTE (2018): "Monetary policy according to HANK," *American Economic Review*, Vol. 108, pp. 697–743.

KEKRE, ROHAN AND MORITZ LENEL (2020): "Monetary Policy, Redistribution and Risk Premia," Working Paper.

KIHLSTROM, RICHARD E AND JEAN-JACQUES LAFFONT (1979): "A general equilibrium entrepreneurial theory of firm formation based on risk aversion," *Journal of Political Economy*, Vol. 87, pp. 719–748.

KINDERMANN, FABIAN AND DIRK KRUEGER (2014): "High marginal tax rates on the top 1%? Lessons from a life cycle model with idiosyncratic income risk,"Technical report, National Bureau of Economic Research.

KITAO, SAGIRI (2008): "Entrepreneurship, taxation and capital investment," *Review of Economic Dynamics*, Vol. 11, pp. 44–69.

KLASS, OREN S, OFER BIHAM, MOSHE LEVY, OFER MALCAI, AND SORIN SOLOMON (2006): "The Forbes 400 and the Pareto wealth distribution," *Economics Letters*, Vol. 90, pp. 290–295.

KLEVMARKEN, N ANDERS, JOSEPH P LUPTON, AND FRANK P STAFFORD (2003): "Wealth Dynamics in the 1980s and 1990s Sweden and the United States," *Journal of Human Resources*, Vol. 38, pp. 322–353.

LOCKWOOD, BENJAMIN B, CHARLES G NATHANSON, AND E GLEN WEYL (2017): "Taxation and the Allocation of Talent," *Journal of Political Economy*, Vol. 125, pp. 1635–1682.

LUCAS JR, ROBERT E (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, pp. 508–523.

LUSARDI, ANNAMARIA, PIERRE-CARL MICHAUD, AND OLIVIA S MITCHELL (2017): "Optimal financial knowledge and wealth inequality," *Journal of Political Economy*, Vol. 125, pp. 431–477.

MAKAROV, DMITRY AND ASTRID V. SCHORNICK (2010): "A Note on Wealth Effect under CARA Utility," *Finance Research Letters*, Vol. 7, pp. 170–177.

MEEUWIS, MAARTEN (2019): "Wealth Fluctuations and Risk Preferences: Evidence from U.S. Investor Portfolios," Job Market Paper Version.

MERTON, ROBERT C (1969): "Lifetime portfolio selection under uncertainty: The continuous-time case," *The review of Economics and Statistics*, pp. 247–257.

MOLL, BENJAMIN (2014): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *The American Economic Review*, Vol. 104, pp. 3186–3221.

MOLL, BENJAMIN AND OLEG ITSKHOKI (2019): "Optimal Development Policies With Financial Frictions," *Econometrica*, Vol. 87, pp. 139–173.

MOSKOWITZ, TOBIAS J AND ANNETTE VISSING-JØRGENSEN (2002): "The returns to entrepreneurial investment: A private equity premium puzzle?" *American Economic Review*, Vol. 92, pp. 745–778.

PENCIAKOVA, VERONIKA (2018): "Diversification & Risky Innovation among US Firms."

PERESS, JOEL (2004): "Wealth, information acquisition, and portfolio choice," *The Review of Financial Studies*, Vol. 17, pp. 879–914.

PEROTTI, ROBERTO (1996): "Growth, income distribution, and democracy: What the data say," *Journal of Economic Growth*, Vol. 1, pp. 149–187.

PFEFFER, FABIAN T, ROBERT F SCHOENI, ARTHUR KENNICKELL, AND PATRICIA ANDRESKI (2016): "Measuring wealth and wealth inequality: Comparing two US surveys," *Journal of economic and social measurement*, Vol. 41, pp. 103–120.

PIKETTY, THOMAS (2018): *Capital in the twenty-first century*: Harvard University Press.

PIKETTY, THOMAS AND EMMANUEL SAEZ (2003): "Income inequality in the United States, 1913–1998," *The Quarterly journal of economics*, Vol. 118, pp. 1–41.

PIKETTY, THOMAS, EMMANUEL SAEZ, AND STEFANIE STANTCHEVA (2014): "Optimal taxation of top labor incomes: A tale of three elasticities," *American economic journal: economic policy*, Vol. 6, pp. 230–71.

QUADRINI, VINCENZO (2000): "Entrepreneurship, saving, and social mobility," *Review of economic dynamics*, Vol. 3, pp. 1–40.

ROBINSON, BAXTER (2012): "Risky business: The choice of entrepreneurial risk under incomplete markets," *Job Market Paper*.

ROTHSCHILD, CASEY AND FLORIAN SCHEUER (2016): "Optimal taxation with rent-seeking," *The Review of Economic Studies*, Vol. 83, pp. 1225–1262.

ROUSSANOV, NIKOLAI (2010): "Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status," *The Journal of Finance*, Vol. 65, pp. 1755–1788.

SAEZ, EMMANUEL (2001): "Using elasticities to derive optimal income tax rates," *The review of economic studies*, Vol. 68, pp. 205–229.

SAEZ, EMMANUEL AND GABRIEL ZUCMAN (2016): "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, Vol. 131, pp. 519–578.

——— (2019): "Progressive wealth taxation," *Brookings Papers on Economic Activity*, Vol. 2019, pp. 437–533.

SAMUELSON, PAUL A (1969): "Lifetime portfolio selection by dynamic stochastic programming," *The review of economics and statistics*, pp. 239–246.

SCHEUER, FLORIAN AND JOEL SLEMROD (2020): "Taxation and the Superrich," *Annual Review of Economics*, Vol. 12, pp. 189–211.

——— (2021): "Taxing our wealth," *Journal of Economic Perspectives*, Vol. 35, pp. 207–30.

SCHULZ, KARL (2021): "Redistribution of Return Inequality."

SHOURIDEH, ALI ET AL. (2012): "Optimal taxation of wealthy individuals," *Work. pap. U. of Pennsylvania*.

SMITH, MATTHEW, DANNY YAGAN, OWEN ZIDAR, AND ERIC ZWICK (2019a): "Capitalists in the Twenty-first Century," *The Quarterly Journal of Economics*, Vol. 134, pp. 1675–1745.

——— (2019b): "Capitalists in the Twenty-First Century*," *The Quarterly Journal of Economics*, Vol. 134, pp. 1675–1745.

STIGLITZ, J. E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica*, Vol. 37, pp. 382–397.

STORESLETTEN, KJETIL, CHRIS I TELMER, AND AMIR YARON (2004): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, Vol. 112, pp. 695–717.

STRAUB, LUDWIG (2018): "Consumption, savings, and the distribution of permanent income," *Unpublished manuscript, Harvard University*.

VERMEULEN, PHILIP (2016): "Estimating the Top Tail of the Wealth Distribution," *The American Economic Review*, Vol. 106, pp. 646–650.

VOITCHOVSKY, SARAH (2005): "Does the profile of income inequality matter for economic growth?" *Journal of Economic Growth*, Vol. 10, pp. 273–296.

XAVIER, INES (2020): "Wealth Inequality in the US: the Role of Heterogeneous Returns."

Appendix

We organize the appendix as follows. Section A contains appendix of the simple analytical model. Section B contains details regarding our empirical work. Section C contains the computational appendix and calibration details.

A Theoretical appendix

A.1 Proofs for Section 2

A.1.1 Final producer maximisation

For clarity, we detail the steps of the final good producer who maximizes the use of labor n and intermediate goods x_s^j, with

$$\max_{\{n, x_s^j\}} \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi} - wn - \sum_s \int_j p_s^j \, x_s^j \, dj \,.$$

Taking the first order condition with respect to labor gives $w = \varphi \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi - 1}$. Plugging this condition together with the assumption that $n = \int_i h^i di = 1$ into the profit function yields

$$\Pi^f = (1 - \varphi) \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi} - \sum_s \int_j p_s^j \, x_s^j dj = (1 - \varphi) \left(\sum_s \int_j x_s^j \, dj \right) - \sum_s \int_j p_s^j \, x_s^j dj \,.$$

A.1.2 Terminal Wealth Distribution

Proof. Because of $u'(c_2^i) > 0$, we know that the second period budget constraint holds in equilibrium with equality. Thus, second period consumption is given by $c_2^i = \underline{r} a_1^i + w + T + R_f a_1^i + \omega_1^i a_1^i (R_f - R_r^i)$. Substituting from the wage rate $w = \varphi Y$ and returns R_f and R_r^i, we get

$$c_2^i = \underline{r} a_1^i + \varphi Y + A(1 - \varphi)(1 - \omega_1^i) a_1^i + \kappa^i (1 - \varphi) \omega_1^i a_1^i + T \,.$$

We obtain that $c_2^i \sim \mathcal{N}\left(\mu_{c_2}^i, \sigma_{c_2}^i\right)$, with

$$\begin{aligned} \mu_{c_2}^i &= \underline{r} a_1^i + \varphi Y + T + A(1 - \varphi)(1 - \omega_1^i) a_1^i + \phi (1 - \varphi) \omega_1^i a_1^i, \\ \sigma_{c_2}^i &= \sigma_\kappa^2 (1 - \varphi)^2 (\omega_1^i a_0^i)^2 \,. \end{aligned}$$

In the Online Appendix OA 1.8, we extend the results with a case in which we introduce labor income risk. $\square$

A.1.3 Proof of Lemma 1

Proof. We first need to derive $\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right]$ analytically. To do so, we use an arbitrary Gaussian distribution with mean $\mu_{c_2}^i$ and variance $(\sigma_{c_2}^i)^2$. Using terminal wealth distribution, we obtain

$$\begin{aligned}
\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right] &= \frac{1}{\alpha_i}\int\left(1-e^{-\alpha_i c_2^i}\right)\times\frac{1}{\sqrt{2\pi(\sigma_{c_2}^i)^2}}e^{-\frac{1}{2(\sigma_{c_2}^i)^2}\left(c_2^i-\mu_{c_2}^i\right)^2}dc_2^i \\
&= \frac{1}{\alpha_i}-\frac{1}{\alpha_i}\int\frac{1}{\sqrt{2\pi(\sigma_{c_2}^i)^2}}e^{-\frac{1}{2(\sigma_{c_2}^i)^2}\left[\left(c_2^i-\mu_{c_2}^i\right)^2+2\alpha_i(\sigma_{c_2}^i)^2c_2^i\right]}dc_2^i \\
&= \frac{1}{\alpha_i}-\frac{1}{\alpha_i}e^{-\alpha_i\mu_{c_2}^i+\frac{1}{2}\alpha_i^2(\sigma_{c_2}^i)^2}\int\frac{1}{\sqrt{2\pi(\sigma_{c_2}^i)^2}}e^{-\frac{1}{2(\sigma_{c_2}^i)^2}\left(c_2^i-(\mu_{c_2}^i-\alpha_i(\sigma_{c_2}^i)^2)\right)^2}dc_2^i
\end{aligned}$$

Recognizing that the term in the integral is the pdf of a normally distributed random variable with mean $\mu_{c_2}^i-\alpha_i(\sigma_{c_2}^i)^2$ and variance $(\sigma_{c_2}^i)^2$, we finally obtain

$$\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right] = \frac{1-e^{-\alpha_i\mu_{c_2}^i+\frac{1}{2}\alpha_i^2(\sigma_{c_2}^i)^2}}{\alpha_i}\,.$$

Under the additional set of assumptions within the special case section, we have $a_1^i \equiv a_0^i$ and $c_2^i = \varphi Y + T + R_f(1-\omega_1^i)a_0^i + R_r^i\omega_1^i a_0^i$ such that $\mu_{c_2}^i = \varphi Y + T + A(1-\varphi)(1-\omega_1^i)a_0^i + \phi(1-\varphi)\omega_1^i a_0^i$ and $\sigma_{c_2}^i = \omega_1^i a_0^i\sigma_\kappa(1-\varphi)$. We solve for the maximization problem given by

$$\max_{\{\omega_1^i\}}\left[1-exp\Big\{-\alpha_i\Big(\mu_{c_2}^i-\frac{\alpha_i}{2}\sigma_{c_2}^i\Big)\Big\}\right]\alpha_i^{-1}\,,$$

Denoting $\mathbb{V} = exp\Big\{-\alpha_i\Big(\mu_{c_2}^i-\frac{\alpha_i}{2}\sigma_{c_2}^i\Big)\Big\}$, the corresponding first order condition is given by

$$-\frac{1}{\alpha_i}\mathbb{V}\left[-\alpha_i(\phi-A)(1-\varphi)a_0^i+\alpha_i^2(a_0^i)^2\omega_1^i\sigma_\kappa^2(1-\varphi)^2\right]=0\,,$$

which results after rearranging

$$\omega_1^i = \frac{\phi-A}{(1-\varphi)\alpha_i\sigma_\kappa^2}(a_0^i)^{-1} = \frac{\phi-A}{(1-\varphi)\sigma_\kappa^2}\frac{\vartheta^i}{\vartheta}(a_0^i)^{\gamma-1}\,.$$

To ensure that the solution is indeed a maximum, we derive the second order condition as

$$-\frac{1}{\alpha_i}\mathbb{V}\left[-\alpha_i(\phi-A)(1-\varphi)a_0^i+\alpha_i^2(a_0^i)^2\omega_1^i\sigma_\kappa^2(1-\varphi)^2\right]^2-\frac{1}{\alpha_i}\mathbb{V}\alpha_i^2\sigma_\kappa^2(1-\varphi)^2(a_0^i)^2<0\,.$$

which completes the proof. ∎

A.1.4 Proof of Proposition 2

Proof. The expression for aggregate innovative asset holdings follows straightforward from integrating over household dynamics while applying the covariance formula

$$\begin{aligned} K_I = (\widetilde{\omega}/\overline{\vartheta})\mathbb{E}\left[\vartheta a_0^{\gamma}\right] &= \frac{\phi - A}{(1-\varphi)\overline{\vartheta}\sigma_\kappa^2}\left(cov(\vartheta, a_0^{\gamma}) + \mathbb{E}\left[\vartheta\right]\mathbb{E}\left[a_0^{\gamma}\right]\right) \\ &= \frac{\phi - A}{(1-\varphi)\overline{\vartheta}\sigma_\kappa^2}\left(\rho_{\vartheta, a_0^{\gamma}}\sigma_\vartheta\sigma_{a_0^{\gamma}} + \mu_\vartheta\mu_{a_0^{\gamma}}\right) . \end{aligned}$$

Aggregate output is given by $Y = \left(\sum_s \int_j x_s^j \, dj\right) n^{\varphi}$, where $n = \int_i e^i \, di = 1$. Given that $K_N + K_I = \mathbb{E}[a_0]$, this can be rewritten after integrating over intermediate goods x_s^j as

$$\begin{aligned} Y &= \int_j x_I^j \, dj + \int_j x_N^j \, dj \\ &= \left[\phi\mu + A(1-\mu)\right]\int_i \omega_1^i a_0^i \, di + A\int_i (1-\omega_1^i)a_0^i \, di \\ &= \mu(\phi - A)K_I + A\mathbb{E}[a_0] = \underbrace{\left[\mu(\phi - A)(K_I/\mathbb{E}[a_0]) + A\right]}_{:=Z}\mathbb{E}[a_0] \end{aligned}$$

The price $\underline{r}$ ensures that total capital distributed to households coincide with the total revenue distributed by the intermediate producer, such that

$$\begin{aligned} \int_i \left(R_f(1-\omega_1^i) + R_r^i\omega_1^i + \underline{r}\right)a_1^i di &= \int_i \left(A(1-\varphi)(1-\omega_1^i) + (\phi\mu + A(1-\mu))(1-\varphi)\omega_1^i\right)a_1^i di \, , \\ \phi(1-\varphi)K_I + \underline{r}\mathbb{E}[a_0] &= (\phi\mu + A(1-\mu))(1-\varphi)K_I \, , \\ \underline{r} &= (\mu - 1)(\phi - A)(1-\varphi)(K_I/\mathbb{E}[a_0]) \, . \end{aligned}$$

where the last equality, regarding the integration of $\int_i R_r^i a_1^i \omega_1^i di$, follows from the simplifying assumption that there is a sub-continuum of households in each state (ϑ^i, a_0^i). □

A.1.5 Proof of Proposition 1

Proof. To prove the result regarding the effect of a mean preserving change in wealth inequality on K_I, we compare the aggregate innovate asset holdings for two economies with different Pareto tails $\eta' \neq \eta$ while keeping aggregate wealth $\mu_{a_0} = \mathbb{E}[a_0]$ constant. We then proceed by case distinction.

Case 1: $\rho_{\vartheta, a_0^{\gamma}} = 0$

The difference in aggregate innovative asset holdings between the two economies is written as

$$\begin{aligned}\Delta^a K_I(\eta',\eta) &= (\widetilde{\omega}/\overline{\vartheta})\mu_\vartheta \left(\frac{\eta'}{\eta'-\gamma}(\underline{a}')^\gamma - \frac{\eta}{\eta-\gamma}\underline{a}^\gamma\right) \\ &= \widetilde{\omega}\frac{\eta}{\eta-\gamma}\underline{a}^\gamma \left(\frac{\eta'}{\eta'-\gamma}\frac{\eta-\gamma}{\eta}\left(\frac{\underline{a}'}{\underline{a}}\right)^\gamma - 1\right) .\end{aligned}$$

Making use of the mean preserving assumption, i.e. $\underline{a}\frac{\eta}{\eta-1} = \underline{a}'\frac{\eta'}{\eta'-1}$, we obtain

$$\Delta^a K_I(\eta',\eta) = \widetilde{\omega}\frac{\eta}{\eta-\gamma}\underline{a}^\gamma \left(\frac{\eta'}{\eta'-\gamma}\frac{\eta-\gamma}{\eta}\left(\frac{\eta'-1}{\eta'}\frac{\eta}{\eta-1}\right)^\gamma - 1\right) .$$

Defining $\chi(\eta,\gamma) \equiv \frac{\eta-\gamma}{\eta}\left(\frac{\eta}{\eta-1}\right)^\gamma$ and taking the derivative of the inner expression w.r.t. η', we get

$$\chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma \left[-\frac{\gamma}{(\eta'-\gamma)^2} + \frac{\gamma}{(\eta'-\gamma)(\eta'-1)}\right] = \chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma \frac{\gamma(1-\gamma)}{(\eta'-\gamma)^2(\eta'-1)} .$$

As a result, we obtain finally

$$\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'} = \widetilde{\omega}\mu_{a_0^\gamma}\chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma \frac{\gamma(1-\gamma)}{(\eta'-\gamma)^2(\eta'-1)} .$$

The Lemma then follows by recognizing that $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'} = 0$ if $\gamma \in \{0,1\}$. Similarly, we obtain $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'} > 0$ if $\gamma \in (0,1)$ and $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'} < 0$ if $\gamma > 1$.

CASE 2: $\rho_{\vartheta,a_0^\gamma} \neq 0$

In the case of an arbitrary correlation between innate risk aversion types and wealth, we obtain

$$\Delta K_I(\eta',\eta) = \Delta^\vartheta K_I(\eta',\eta) + \Delta^a K_I(\eta',\eta) ,$$

where the first term denotes distributional relevance arising from the selection effect, whereas the second term resembles distributional relevance arising from wealth dependent risk taking. Notice that the latter is equivalent to Case 1. Contrary, the first effect can be written as

$$\Delta^\vartheta K_I(\eta',\eta) = (\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\frac{\rho_{\vartheta,a_0^\gamma}(\eta',\underline{a}',\cdot)}{\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}\frac{\sigma_{a_0^\gamma}(\eta',\underline{a}')}{\sigma_{a_0^\gamma}(\eta,\underline{a})} - 1\right) .$$

Using the relation $\underline{a}\frac{\eta}{\eta-1} = \underline{a}'\frac{\eta'}{\eta'-1}$, a change in the Pareto tail η' preserves the mean wealth if $\underline{a}'(\eta') = \underline{a}\left(\frac{\eta}{\eta-1}\right)\left(\frac{\eta'-1}{\eta'}\right)$. Using a first order Taylor approximation of $\rho_{\vartheta,a_0^\gamma}(\eta',\underline{a}'(\eta'),\cdot)$ around η, we obtain:

$$\rho_{\vartheta,a_0^\gamma}(\eta',\underline{a}'(\eta'),\cdot) \approx \rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot) + \left|\frac{\partial\rho^1_{\vartheta,a_0^\gamma}(\eta',\underline{a}'(\eta'),\cdot)}{\partial\eta} + \frac{\partial\rho^2_{\vartheta,a_0^\gamma}(\eta',\underline{a}'(\eta'),\cdot)}{\partial\underline{a}'(\eta')}\frac{\partial\underline{a}'(\eta')}{\partial\eta'}\right|_{\eta'=\eta}(\eta'-\eta)$$

$$\approx \rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot) + \left(\frac{\partial\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial\eta} + \frac{\partial\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial\underline{a}}\frac{\underline{a}}{\eta(\eta-1)}\right)(\eta'-\eta)\ .$$

Substituting the previous expression into the one for $\Delta^\vartheta K_I(\eta',\eta)$ we arrive at

$$\Delta^\vartheta K_I(\eta',\eta) \approx (\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\left[1+\frac{\left(\frac{\partial\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial\eta}+\frac{\partial\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial\underline{a}}\frac{\underline{a}}{\eta(\eta-1)}\right)}{\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}(\eta'-\eta)\right]\frac{\sigma_{a_0^\gamma}(\eta',\underline{a}')}{\sigma_{a_0^\gamma}(\eta,\underline{a})}-1\right)\ .$$

With a slight abuse of notation, we can take the derivative w.r.t. η' to obtain

$$\frac{\partial\Delta^\vartheta K_I(\eta',\eta)}{\partial\eta'} \approx (\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\frac{1}{\sigma_{a_0^\gamma}}\frac{\partial\sigma'_{a_0^\gamma}}{\partial\eta'}+\frac{\frac{\partial\rho_{\vartheta,a_0^\gamma}}{\partial\eta}+\frac{\partial\rho_{\vartheta,a_0^\gamma}}{\partial\underline{a}}\frac{\underline{a}}{\eta(\eta-1)}}{\rho_{\vartheta,a_0^\gamma}}\left(\frac{\sigma'_{a_0^\gamma}}{\sigma_{a_0^\gamma}}+(\eta'-\eta)\frac{1}{\sigma_{a_0^\gamma}}\frac{\partial\sigma'_{a_0^\gamma}}{\partial\eta'}\right)\right)\ .$$

To get an impression about the sign of the previous derivative, let us first analyze the sign of $\frac{\partial\sigma'_{a_0^\gamma}}{\partial\eta'}$. Notice that it is straightforward to show that a_0^γ follows a $\mathcal{P}a(\underline{a}^\gamma,\frac{\eta}{\gamma})$ distribution. As a result, the variance is given by

$$\sigma'_{a_0^\gamma} = (\underline{a}')^{2\gamma}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma}-1\right)^2\left(\frac{\eta'}{\gamma}-2\right)} = \left(\underline{a}\frac{\eta}{\eta-1}\right)^{2\gamma}\left(\frac{\eta'-1}{\eta'}\right)^{2\gamma}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma}-1\right)^2\left(\frac{\eta'}{\gamma}-2\right)}\ ,$$

where again the last equality follows from using $\underline{a}\frac{\eta}{\eta-1} = \underline{a}'\frac{\eta'}{\eta'-1}$. Defining the auxiliary variable $\tilde{\chi}(\eta,\gamma,\underline{a}) \equiv \left(\underline{a}\frac{\eta}{\eta-1}\right)^{2\gamma}$, we obtain:

$$\frac{\partial\sigma'_{a_0^\gamma}}{\partial\eta'} = \tilde{\chi}(\eta,\gamma,\underline{a})\left(2\gamma\left(\frac{\eta'-1}{\eta'}\right)^{2\gamma-1}\frac{1}{(\eta')^2}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma}-1\right)^2\left(\frac{\eta'}{\gamma}-2\right)}\right)$$

$$+\tilde{\chi}(\eta,\gamma,\underline{a})\left(\frac{\eta'-1}{\eta'}\right)^{2\gamma}\left(\frac{\frac{1}{\gamma}\left(\frac{\eta'}{\gamma}-1\right)^2\left(\frac{\eta'}{\gamma}-2\right)-\frac{\eta'}{\gamma}\left[\frac{2}{\gamma}\left(\frac{\eta'}{\gamma}-1\right)\left(\frac{\eta'}{\gamma}-2\right)+\frac{1}{\gamma}\left(\frac{\eta'}{\gamma}-1\right)^2\right]}{\left(\frac{\eta'}{\gamma}-1\right)^4\left(\frac{\eta'}{\gamma}-2\right)^2}\right)\ .$$

Collecting terms leads to

$$\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} = \tilde{\chi}(\eta,\gamma,\underline{a}) \left(\frac{\eta'-1}{\eta'}\right)^{2\gamma} \frac{1}{\left(\frac{\eta'}{\gamma}-1\right)^2 \left(\frac{\eta'}{\gamma}-2\right)} \left[\frac{2}{\eta'-1} + \frac{1}{\gamma} - \frac{2\eta'}{\gamma(\eta'-\gamma)} - \frac{\eta'}{\gamma(\eta'-2\gamma)}\right]$$
$$= \frac{1}{\eta'}\sigma'_{a_0^\gamma}\left[1 + \frac{2\gamma}{\eta'-1} - \frac{2\eta'}{(\eta'-\gamma)} - \frac{\eta'}{(\eta'-2\gamma)}\right] .$$

In order to determine the sign of the bracket term, one can simplify to

$$\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} = \frac{2}{\eta'}\sigma'_{a_0^\gamma}\left[\frac{\gamma(1-2\gamma)(\eta'-\gamma) - \eta'(\eta'-1)(\eta'-2\gamma)}{(\eta'-1)(\eta'-\gamma)(\eta'-2\gamma)}\right] .$$

It is straightforward to show that the previous term is (weakly) negative if

$$\eta' \geq 2\gamma + \gamma\frac{(1-2\gamma)(\eta'-\gamma)}{\eta'(\eta'-1)} .$$

The left hand side of this expression is increasing in η', whereas the right hand side is decreasing if $\gamma \leq \frac{1}{2}$ and increasing if $\gamma > \frac{1}{2}$. Hence, for the case of $\gamma \leq \frac{1}{2}$, we obtain after substituting $\eta' = 2\gamma$ an upper limit of the right hand side given by $\overline{\eta} = \frac{3}{2}\gamma$. Contrary, in the case of $\gamma > \frac{1}{2}$ a straightforward application of L'Hopital's rule results in $\overline{\eta} = 2\gamma$. As a result, we obtain that $\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} \leq 0 \; \forall \eta' \geq \overline{\eta} = \max\{\frac{3}{2}\gamma, 2\gamma\} = 2\gamma$, which trivially holds due to the implicit assumed finite variance of the $\mathcal{P}a(\underline{a}^\gamma, \frac{\eta}{\gamma})$ distribution. Consequently, the result of proposition 1 follows (given a small change in the wealth Pareto tail). □

A.1.6 Proof of Proposition 1

Proof. The Farlie-Gumbel-Morgenstern (FGM) copula can be written for two arbitrary cumulative distribution functions $\{F(x_1), F(x_2)\}$ as

$$F(x_1,x_2) = C^{FGM}(F(x_1),F(x_2)) = F(x_1)F(x_2) + \varrho F(x_1)F(x_2)(1-F(x_1))(1-F(x_2)) ,$$

where $\varrho \in [-1,1]$. The joint probability density function of $f(x_1,x_2)$ is the obtained by

$$f(x_1,x_2) = (1+\varrho\,(1-2F(x_1))\,(1-2F(x_2)))\,f(x_1)f(x_2)$$
$$= (1+\varrho+2\varrho\,(2F(x_1)F(x_2) - F(x_1) - F(x_2)))\,f(x_1)f(x_2) .$$

Under this assumption that $\vartheta \sim \mathcal{P}a(\underline{\vartheta}, \epsilon)$ and $a_0 \sim \mathcal{P}a(\underline{a}, \eta)$ this provides us with

$$f(\vartheta, a_0) = (1+\varrho)f(\vartheta)f(a_0) + 2\varrho\left[2\left(\frac{\underline{\vartheta}}{\vartheta}\right)^\epsilon \left(\frac{\underline{a}}{a_0}\right)^\eta - \left(\frac{\underline{\vartheta}}{\vartheta}\right)^\epsilon - \left(\frac{\underline{a}}{a_0}\right)^\eta\right] f(\vartheta)f(a_0) .$$

where the marginals are given by $f(\vartheta)$ and $f(a_0)$. Given the Pareto assumptions, we have $\mu_\vartheta \equiv \overline{\vartheta} = \underline{\vartheta}\frac{\epsilon}{\epsilon-1}$ and $\mu_{a_0^\gamma} = \underline{a}^\gamma \frac{\eta}{\eta-\gamma}$. In order to derive $cov(\vartheta, a_0^\gamma)$, we need to compute $\mathbb{E}\left[\vartheta a_0^\gamma\right]$:

$$\mathbb{E}\left[\vartheta a_0^\gamma\right] = \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta a_0^\gamma f(\vartheta, a_0) d\vartheta da_0 \, .$$

Using the FGM copula, we proceed in four steps:

$$\begin{aligned}
(1+\varrho)\underline{\vartheta}^\epsilon \underline{a}^\eta \epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-\epsilon} a_0^{\gamma-\eta-1} d\vartheta da_0 &= (1+\varrho)\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma \frac{\eta}{\eta-\gamma} \, , \\
4\varrho\underline{\vartheta}^{2\epsilon}\underline{a}^{2\eta}\epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-2\epsilon} a_0^{\gamma-2\eta-1} d\vartheta da_0 &= 4\varrho\underline{\vartheta}\frac{\epsilon}{2\epsilon-1}\underline{a}^\gamma \frac{\eta}{2\eta-\gamma} \, , \\
-2\varrho\underline{\vartheta}^{2\epsilon}\underline{a}^{\eta}\epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-2\epsilon} a_0^{\gamma-\eta-1} d\vartheta da_0 &= -2\varrho\underline{\vartheta}\frac{\epsilon}{2\epsilon-1}\underline{a}^\gamma \frac{\eta}{\eta-\gamma} \, , \\
-2\varrho\underline{\vartheta}^{\epsilon}\underline{a}^{2\eta}\epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-\epsilon} a_0^{\gamma-2\eta-1} d\vartheta da_0 &= -2\varrho\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma \frac{\eta}{2\eta-\gamma} \, .
\end{aligned}$$

Combining the previous four equations gives

$$\begin{aligned}
cov(\vartheta, a_0^\gamma) &= \mathbb{E}\left[\vartheta a_0^\gamma\right] - \mathbb{E}\left[\vartheta\right]\mathbb{E}\left[a_0^\gamma\right] \\
&= \underline{\vartheta a}^\gamma \left[\varrho\frac{\epsilon}{\epsilon-1}\frac{\eta}{\eta-\gamma} + 4\varrho\frac{\epsilon}{2\epsilon-1}\frac{\eta}{2\eta-\gamma} - 2\varrho\frac{\epsilon}{2\epsilon-1}\frac{\eta}{\eta-\gamma} - 2\varrho\frac{\epsilon}{\epsilon-1}\frac{\eta}{2\eta-\gamma}\right]
\end{aligned}$$

Further simplifications result in

$$cov(\vartheta, a_0^\gamma) = \underline{\vartheta a}^\gamma \varrho \frac{\epsilon}{(\epsilon-1)(2\epsilon-1)} \frac{\eta\gamma}{(\eta-\gamma)(2\eta-\gamma)} \, .$$

As a result, aggregate innovative asset holdings from Lemma 2 are given by

$$K_I = (\widetilde{\omega}/\overline{\vartheta})\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma\frac{\eta}{\eta-\gamma} \, .$$

Aggregate risk free capital holdings from Lemma 1 are (weakly) positive if the condition $\mu_{a_0} \geq K_I$ holds, which can be rewritten as

$$\frac{\eta-\gamma}{\eta-1} \geq (\widetilde{\omega}/\overline{\vartheta})\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\epsilon}{\epsilon-1}\underline{\vartheta a}^{\gamma-1} \, .$$

Finally, as $\mu_\vartheta = \mathbb{E}[\vartheta] = \overline{\vartheta}$, let $\tilde{\omega} = (\widetilde{\omega}/\overline{\vartheta})\mu_\vartheta \underline{a}^{\gamma-1} = \widetilde{\omega}\underline{a}^{\gamma-1}$, $B = -\mu\tilde{\omega}(\phi - A)$ and $C = (2\epsilon - 1)$, we derive the marginal effect of a change in the Pareto tail η on wealth-normalized output $\widetilde{Y}(\eta) \equiv$

$\frac{Y(\eta)}{\mu_{a_0}} = A + \mu\widetilde{\omega}(\phi - A)\Psi(\eta)$ with $\Psi(\eta) = \left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\eta-1}{\eta-\gamma}$ as:

$$
\begin{aligned}
\frac{\partial\widetilde{Y}(\eta)}{\partial\eta} &= -B\frac{\partial\Psi(\eta)}{\partial\eta} \\
&= -B\left[\left(\frac{1}{\eta-\gamma}\right)\left(1+\frac{\varrho\gamma}{C(2\eta-\gamma)}\right) - \left(\frac{\eta-1}{(\eta-\gamma)^2}\right)\left(1+\frac{\varrho\gamma}{C(2\eta-\gamma)}\right) - 2\left(\frac{\varrho\gamma}{C(2\eta-\gamma)^2}\right)\left(\frac{\eta-1}{\eta-\gamma}\right)\right] \\
&= B\left[(\gamma-1)\underbrace{\left(\frac{1}{(\eta-\gamma)^2}\right)}_{:=\Omega^{\gamma}} + \varrho(\gamma-1)\underbrace{\left(\frac{(\gamma(2\eta-\gamma)+2(\eta-\gamma)(\eta-1))}{C(2\eta-\gamma)^2(\eta-\gamma)^2}\right)}_{:=\Omega^{\varrho\gamma}} + \varrho\underbrace{\left(\frac{2(\eta-1)}{C(2\eta-\gamma)^2(\eta-\gamma)}\right)}_{:=\Omega^{\varrho}}\right] \\
&= B\left(\frac{1}{C(2\eta-\gamma)^2(\eta-\gamma)^2}\right)\left[(\gamma-1)C(2\eta-\gamma)^2 + \varrho(\gamma-1)2(\eta-1)(\eta-\gamma) + \varrho(2\eta(\eta-1)+\gamma)\right]
\end{aligned}
$$

where the before last line follows from $\varrho\gamma = \varrho + \varrho(\gamma-1)$. In order to determine the sign of the derivatives, we need to know the sign of $\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)$. To do so, let us assume that

$$1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)} \leq 0 \Leftrightarrow 1 \leq -\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)} \leq \frac{\gamma}{(2\epsilon-1)(2\eta-\gamma)},$$

where the last inequality follows from $\varrho \in [-1,1]$. As we have $\epsilon > 1$ and $\eta > \gamma$, an upper bound of $\frac{\gamma}{(2\epsilon-1)(2\eta-\gamma)}$ is given by $\frac{\gamma}{\epsilon\eta}$, which is strictly smaller than one. Hence, we obtain a contradiction and conclude that $1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}$ is strictly positive. This completes the proof of Proposition 1. □

A.1.7 Existence of the Growth Irrelevance Frontier of Wealth Inequality

Lemma 6 (EXISTENCE GIF). *For a tail of wealth η and of type ϵ, type dependence ϱ, and wealth-dependent risk taking $\gamma \in (\underline{\gamma}, \overline{\gamma})$ with $\underline{\gamma} = \frac{2(2\epsilon-1)}{1+2(2\epsilon-1)}$ and $\overline{\gamma} = 2$, there exists a η^* which lies on the GIF.*

(a) For $\gamma \in (1, \overline{\gamma})$ and $-1 \leq \varrho \leq 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma}$, the GIF exists for a unique $\eta^* \in (\gamma, \infty)$.

(b) With $\gamma = 1$ and $\varrho = 0$, any Pareto tail $\eta^* \in (\gamma, \infty)$ lies on the GIF.

(c) For $\gamma \in (\underline{\gamma}, 1)$, and $2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \leq \varrho \leq 1$, the GIF exists for a unique $\eta^* \in (\underline{\eta}^{dc}, \infty)$, $\underline{\eta}^{dc} > 1$.

item a. proves that a negative *type* dependence ($\varrho < 0$) is needed for the existence of an economy which lies on the GIF if the *scale* dependence is positive ($\gamma > 0$). Conditions in item b. without *type* dependence requires no *scale* dependence for an economy to lie on the GIF, and conditions in item c. with positive *type* dependence shows that an economy with some $\gamma < 0$ can be located on the GIF.

A.1.8 Proof of Lemma 6: Existence of the $isoG(\eta, 0)$

Proof. Using Lemma 1, the *iso-growth* at level $\overline{g}$ can be implicitly defined as

$$(\phi - A)\overline{g} = -(\phi - A)\Big[(\gamma-1)\Omega^{\gamma} + \varrho(\gamma-1)\Omega^{\varrho\gamma} + \varrho\Omega^{\varrho}\Big] \tag{33}$$

Which, for a level $\overline{g} = 0$ can be rewritten as:

$$\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{1-\gamma}{(\eta-\gamma)^2} - 2\frac{\eta-1}{\eta-\gamma}\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)^2} = 0\,,$$

which we can rearrange to

$$\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)^2(\eta-\gamma)^2}\Big((1-\gamma)(2\eta-\gamma) - 2(\eta-1)(\eta-\gamma)\Big) = -\frac{1-\gamma}{(\eta-\gamma)^2}\,,$$

If $(1-\gamma)(2\eta-\gamma) - 2(\eta-1)(\eta-\gamma) \neq 0$ (which only occurs in the case of $\gamma < 1$) we can state the GIF algebraically as

$$\varrho(\gamma, \eta, \epsilon; \overline{g} = 0) = \frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2\eta-\gamma)^2}{2(\eta-1)(\eta-\gamma) - (1-\gamma)(2\eta-\gamma)}\,. \tag{34}$$

The sign of the derivative w.r.t. to the Pareto tail is determined by

$$\begin{aligned}\operatorname{sgn}\left(\frac{\partial\varrho}{\partial\eta}\right) &= \operatorname{sgn}\Big((1-\gamma)\Big)\operatorname{sgn}\Big(4(2\eta-\gamma)\Big(2(\eta-1)(\eta-\gamma) - (1-\gamma)(2\eta-\gamma)\Big) - (2\eta-\gamma)^2\Big(4(\eta-1)\Big)\Big)\\ &= \operatorname{sgn}\Big((1-\gamma)\Big)\operatorname{sgn}\Big((\gamma-\eta)(2-\gamma)\Big)\,.\end{aligned}$$

Hence, we obtain due to $\eta > \gamma$

$$\frac{\partial\varrho}{\partial\eta}\begin{cases} < 0 & \text{if } \gamma > 2\,,\\ = 0 & \text{if } \gamma = 2\,,\\ > 0 & \text{if } 1 < \gamma < 2\,,\\ = 0 & \text{if } \gamma = 1\,,\\ < 0 & \text{if } 0 < \gamma < 1\,.\end{cases}$$

Before turning to the existence of η^{GIF}, we first study the limits of (34). Thus, we obtain

$$\lim_{\eta\to\gamma^+}\varrho(\gamma,\eta,\epsilon) = -\frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2\eta-\gamma)}{1-\gamma} = -(2\epsilon-1)\,.$$

Similarly, we have

$$\lim_{\eta\to 1^+}\varrho(\gamma,\eta,\epsilon) = -\frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2-\gamma)}{1-\gamma} = -\frac{(2\epsilon-1)(2-\gamma)}{\gamma}\,.$$

Finally, by an application of L'Hopitals rule we derive

$$\lim_{\eta\to\infty} \varrho(\gamma,\eta,\epsilon) = \frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{2\eta-\gamma}{\eta-1}|_{\eta=\infty} = 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} .$$

As a result, we obtain the following bounds on the copula dependence parameter

$$\begin{aligned} -(2\epsilon-1) < \varrho < 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \qquad & \text{if } \gamma > 1 \\ \varrho = 0 \qquad & \text{if } \gamma = 1 \\ -\frac{(2\epsilon-1)(2-\gamma)}{\gamma} < \varrho < 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \qquad & \text{if } \gamma < 1 \end{aligned}$$

It is straightforward to see for $\gamma > 2$ that the required $\varrho \notin \mathcal{R}$ such that the GIF is empty (i.e. $\nexists\ \eta^* \in (\gamma,\infty)$ s.t. $\text{GIF}(\eta^*) = 0$). Contrary, for $1 < \gamma < 2$ there exists by an application of the intermediate value theorem a unique $\eta^* \in (\gamma,\infty)$ such that $\text{GIF}(\eta^*) = 0$ if $-1 < \varrho < \overline{\rho}^{FGM} < 0$, where $\overline{\rho}^{FGM} \equiv 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma}$. Additionally, in the case of $\gamma = 1$, being on the growth irrelevance frontier requires $\varrho = 0$. As a result, for any $\eta \in (1,\infty)$ the GIF goes through the point $\{\gamma = 1, \varrho = 0\}$. Finally, in the case of $\gamma < 1$ we can show that the growth irrelevance frontier is discontinuous at the points (cf. denominator of equation (34))

$$\eta_{dc}^{1,2} = 1 \pm \sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)} .$$

Recognizing that $3\gamma - \gamma^2$ is strictly increasing in γ on the interval $(0,1)$, we conclude that the expression in the square brackets is strictly positive and lies on the interval $(0,1)$. Due to the imposition of $\eta > 1$, we can hence exclude the smaller solution. Let us subsequently denote by $\eta_{dc}^* \in (1,2)$ the only feasible discontinuity point. As $\frac{\partial\varrho}{\partial\eta} < 0$, the denominator of (34) is strictly increasing in η and $\lim\limits_{\eta\to 1^+} < -1$ for $\gamma \in (0,1)$, we conclude that $\varrho > 0$ is a necessary condition for the existence of a GIF solution. This implies that the feasible set of Pareto tails reduces to $\eta \in (\eta_{dc}^*, \infty)$. Ensuring that $\varrho \leq 1$ gives us finally a lower bound on the wealth dependent risk taking parameter such that $\gamma \geq \underline{\gamma} = \frac{2(2\epsilon-1)}{1+2(2\epsilon-1)}$. As a result, for all $\gamma < \underline{\gamma}$ a GIF solution does not exists. Contrary, by an application of the intermediate value theorem an unique $\eta^* \in (\eta_{dc}^*, \infty)$ exists for $\underline{\gamma} \leq \gamma < 1$. This completes the proof of Lemma 6. □

A.1.9 Proof of Lemma 3: Properties of the GIF and iso-growth

Proof. Using equation (34) from Lemma 6, we obtain

$$\frac{\partial \varrho}{\partial \gamma} = -(2\epsilon-1)\frac{1}{\gamma^2}\frac{(2\eta-\gamma)^2}{2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)}$$
$$+(2\epsilon-1)\frac{1-\gamma}{\gamma}\frac{-2(2\eta-\gamma)\left[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\right]-(2\eta-\gamma)^2\left[2(1-\gamma)+1\right]}{\Big(2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big)^2}\,.$$

Hence, the sign of the previous expression is determined by the sign of

$$\operatorname{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big) = \Big(-(2\eta-\gamma)^2\Big[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big]\Big)$$
$$+\Big((1-\gamma)\gamma(2\eta-\gamma)\Big[-4(\eta-1)(\eta-\gamma)+2(1-\gamma)(2\eta-\gamma)-2(2\eta-\gamma)(1-\gamma)-2(\eta-\gamma)\Big]\Big)\,.$$

Simplifying terms provides us with

$$\operatorname{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big) = \operatorname{sgn}\Big(\underbrace{-(2\eta-\gamma)\Big[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big]}_{\equiv\mathcal{A}}+\underbrace{(1-\gamma)\gamma\Big[-4(\eta-1)(\eta-\gamma)-2(\eta-\gamma)\Big]}_{\equiv\mathcal{B}}\Big)\,.$$

Let us begin with the case $\gamma = 1$. It is straightforward to see that

$$\operatorname{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big)|_{\gamma=1} = \operatorname{sgn}(\mathcal{A}) = \operatorname{sgn}\Big(-2(2\eta-\gamma)(\eta-1)(\eta-\gamma)\Big) < 0\,,$$

such that the GIF is strictly decreasing in the point $\gamma = 1$. For the case $\underline{\gamma} < \gamma < 1$, the reasoning is slightly more evolved. First, notice that $\mathcal{B} < 0$ in this case. Second, requiring that the inner bracket of $\mathcal{A}$ is weakly positive is equivalent to requiring that

$$2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma) \geq 0 \Leftrightarrow 2\left(\eta^2-\eta\gamma-\eta+\gamma\right)-2\eta+\gamma+2\eta\gamma-\gamma^2 \geq 0$$

Collecting terms gives us the following condition

$$2\eta^2-4\eta+3\gamma-\gamma^2 \geq 0\,. \tag{35}$$

We know from Lemma 6 that the GIF is only defined in this case if $\eta > \eta^*_{dc} = 1+\sqrt{1-\frac{1}{2}\left(3\gamma-\gamma^2\right)}$. Substituting this expression into the former inequality yields

$$\begin{aligned}\text{LHS} &= 2\left(1+\sqrt{1-\frac{1}{2}\left(3\gamma-\gamma^2\right)}\right)^2-4\left(1+\sqrt{1-\frac{1}{2}\left(3\gamma-\gamma^2\right)}\right)+3\gamma-\gamma^2\\ &= 2+4\sqrt{1-\frac{1}{2}\left(3\gamma-\gamma^2\right)}+2\left(1-\frac{1}{2}\left(3\gamma-\gamma^2\right)\right)-4-4\sqrt{1-\frac{1}{2}\left(3\gamma-\gamma^2\right)}+3\gamma-\gamma^2\\ &= 0\,.\end{aligned}$$

As the left hand side of the inequality (35) is strictly increasing in η on the set $\eta > \eta^*_{dc} > 1$, we know that the above inequality is always strictly satisfied. Hence, we conclude that $\mathcal{A} < 0$. Finally, this provides us with

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big)|_{\underline{\gamma} \leq \gamma < 1} = \text{sgn}\,(\mathcal{A} + \mathcal{B}) < 0\;.$$

Let us finally consider the case $1 \leq \gamma < \overline{\gamma}$. It is evident that $\text{sgn}(\mathcal{A}) < 0$ and $\text{sgn}(\mathcal{B}) > 0$ hold in this case. Hence, the sign of the derivative is *a priori* undetermined. To show our claim, let us first rewrite the claim on the sign of our initial inequality as

$$\begin{aligned} &-(2\eta-\gamma)\Big[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big]+(1-\gamma)\gamma\Big[-4(\eta-1)(\eta-\gamma)-2(\eta-\gamma)\Big]<0 \\ &\Leftrightarrow (2\eta-\gamma)(\eta+\gamma-2)+2(\eta-1)\gamma(1-\gamma)>0\;. \end{aligned} \tag{36}$$

We know that $\eta > \gamma$ has to hold. Substituting $\eta = \gamma$ into the previous inequality yields

$$2\gamma(\gamma-1) > 2\gamma(\gamma-1)^2\;,$$

which is trivially satisfied for $\gamma < 2$. Hence, it suffices to show that the left hand side of (36) is increasing in η. To do so, let us take the derivative of (36) w.r.t. η and let us simultaneously impose that the derivative is positive:

$$\mathcal{Q}(\eta) \equiv 2(\eta+\gamma-2)+2\eta-\gamma+2\gamma(1-\gamma) = 4(\eta-1)+3\gamma-2\gamma^2 > 4(\gamma-1)+3\gamma-2\gamma^2 \equiv \mathcal{Q}(\gamma),$$

where the second last inequality holds due to $\eta > \gamma$. Hence, $\mathcal{Q}(\gamma) > 0$ implies also $4(\eta-1)+3\gamma-2\gamma^2 > 0$. To show the validity of the previous inequality, let us compute the values of γ of the second order polynomial $\mathcal{Q}(\gamma)$ for which the function equals exactly zero: $\tilde{\gamma}^{1,2} = \frac{-7\pm\sqrt{49-32}}{-4} = \frac{7}{4} \pm \frac{1}{4}\sqrt{17}$. As a consequence, we have that $\tilde{\gamma}^1 < 1$ and $\tilde{\gamma}^2 > 2$ such that $\mathcal{Q}(\gamma)$ is due to continuity strictly positive on the interval $\gamma \in [1,2]$. As a result, we know that $\mathcal{Q}(\eta)$ is also strictly positive on the entire interval $\gamma \in [1,2]$ which proves that equation (36) is satisfied. This shows

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big)|_{1 < \gamma < \overline{\gamma}} = \text{sgn}\,(\mathcal{A} + \mathcal{B}) < 0\;,$$

such that we overall obtain

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big)|_{\underline{\gamma} \leq \gamma < \overline{\gamma}} = \text{sgn}\,(\mathcal{A} + \mathcal{B}) < 0\;.$$

Let us finally define the growth irrelevance equation (34) by $\varrho \equiv \mathcal{G}(\gamma)$, where $\{\epsilon, \eta\}$ enter the $\mathcal{G}$ function as constants. Recognize that $\mathcal{G}$ is strictly decreasing on the interval $\underline{\gamma} \leq \gamma < \overline{\gamma}$ and thus *injective*. Additionally, it is differentiable at $\mathcal{G}^{-1}(\varrho)$ and hence continuous on the interval $\mathbb{G}$. As a result, we can define the inverse function of the growth irrelevance frontier equation (34) by

$\gamma \equiv \mathcal{G}^{-1}(\varrho)$. Consequently, it is straightforward to obtain

$$\frac{\partial \gamma}{\partial \varrho} = \frac{\partial \mathcal{G}^{-1}(\varrho)}{\partial \varrho} = \frac{1}{\mathcal{G}'(\mathcal{G}^{-1}(\varrho))} = \frac{1}{\mathcal{G}'(\gamma)} < 0 \,,$$

which completes the first part of the proof of Lemma 3.

PART 2. Let us consider now the general growth irrelevance frontier at an arbitrary growth level $\overline{g}$, possibly different from zero. Without loss of generality let us further assume that $\underline{a} = 1$. The GIF is then implicitly characterized by

$$\frac{(1-\gamma)(2\eta-\gamma) - 2(\eta-1)(\eta-\gamma)}{(2\epsilon-1)(2\eta-\gamma)^2(\eta-\gamma)^2}\gamma\varrho + \frac{1-\gamma}{(\eta-\gamma)^2} = \chi_{\overline{g}}\overline{g} \,,$$

where $\chi_{\overline{g}}$ denotes a strictly positive constant which is independent of $\{\varrho, \gamma\}$. Total differentiation of the previous equation yields

$$\chi_{\rho} d\varrho + \chi_{\gamma} d\gamma = \chi_{\overline{g}} d\overline{g} \,.$$

Rearranging the previous condition results in

$$d\gamma = -\frac{\chi_{\rho}}{\chi_{\gamma}} d\varrho + \frac{\chi_{\overline{g}}}{\chi_{\gamma}} d\overline{g} \,.$$

On the restricted set of Lemma 6, we have $\chi_{\rho} < 0$. Additionally, we have that $\frac{\chi_{\rho}}{\chi_{\gamma}} = -\frac{1}{\mathcal{G}'(\gamma)}$, which implies $\chi_{\gamma} < 0$. Hence, an increase in $\overline{g}$ (i.e. lower growth rate) decreases γ, conditional on ϱ. As a result, the GIF shifts downwards. Similarly, if $\overline{g}$ decreases (i.e. higher growth rate), γ increases conditional on ϱ which shifts the GIF upwards. This concludes the proof. $\square$

A.1.10 Proof of Lemma 4

We first solve for the consumption equivalent variation $\Delta^{CE,i}$, defined as the amount of consumption that makes an individual indifferent between the reformed economy with progressivity p_a and the initial status quo situation, such that $\mathbb{E}[u(\widetilde{c}_2^i - \Delta^{CE,i})] = \mathbb{E}[u(c_2^i)]$. We get:

$$E[(1 - exp(-\widetilde{\alpha}^i(\widetilde{c}_2^i - \Delta^{CE,i})))]/\widetilde{\alpha}^i = E[(1 - exp(-\alpha^i c_2^i))]/\alpha^i \,,$$

which is, under the generalized CARA and $\Delta^{\alpha} = \left(1/\widetilde{\alpha}^i - 1/\alpha^i\right)$, equivalent to

$$\begin{aligned}
\Delta^{\alpha} - exp\left(-\widetilde{\alpha}^i(\widetilde{x}_2^i - \Delta^{CE,i})\right)/\widetilde{\alpha}^i &= -exp\left(-\alpha^i x_2^i\right)/\alpha^i \\
exp\left(-\widetilde{\alpha}^i(x_2^i - \Delta^{CE,i}) + \alpha^i x_2^i\right) &= \left[1 + \Delta^{\alpha} exp\left(\alpha^i x_2^i\right)\alpha^i\right](\widetilde{\alpha}^i/\alpha^i) \\
-\widetilde{\alpha}^i(\widetilde{x}_2^i - \Delta^{CE,i}) + \alpha^i x_2^i &= \Delta^c \,,
\end{aligned}$$

where $\widetilde{x}_2^i$ and x_2^i denote the certainty equivalents. Rearranging terms, this yields

$$\Delta^{CE,i} = \widetilde{x}_2^i - x_2^i + \frac{\Delta^c}{\widetilde{\alpha}^i},$$

with $\Delta^c = -\widetilde{\alpha}^i \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) x_2^i + \ln\left(1 + \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) exp\left(\alpha^i x_2^i\right)\right) + \ln\left(\frac{\widetilde{\alpha}^i}{\alpha^i}\right)$, a term that arises because a change in the progressivity p_a affects a_0^i which impacts the curvature of the utility function through a change in the risk aversion $\widetilde{\alpha}^i$.

We now analyze the effects of introducing a proportional tax on each component. First, let us analyze the effect on $\widetilde{x}_2^i = \widetilde{\mu}_2^i - (\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2$, with

$$\begin{aligned} \widetilde{\mu}_2^i &= \varphi Y + T + \underline{r}\widetilde{a}_0^i + A(1-\varphi)\widetilde{a}_0^i + (\phi - A)(1-\varphi)\omega_1^i \widetilde{a}_0^i \\ ((\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2) &= \frac{1}{2}\sigma_\kappa^2 (\widetilde{a}_0^i)^\gamma \widetilde{\omega}^2 \frac{\vartheta^i}{\overline{\vartheta}}(1-\varphi)^2, \end{aligned}$$

we get

$$\begin{aligned} \frac{\partial \widetilde{\mu}_2^i}{\partial p_a} &= \varphi \frac{\partial Y}{\partial \widetilde{\eta}}\frac{\partial \widetilde{\eta}}{\partial p_a} + \frac{\partial \underline{r}}{\partial \widetilde{\eta}}\frac{\partial \widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i + \frac{\partial T}{\partial p_a} + \frac{\partial \widetilde{a}_0{}^i}{\partial p_a}\left[A(1-\varphi) + \underline{r} + \gamma \widetilde{\omega}\frac{\vartheta^i}{\overline{\vartheta}}(\widetilde{a}_0^i)^{\gamma-1}(1-\varphi)(\phi - A)\right] \\ \frac{\partial((\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2)}{\partial p_a} &= \frac{\partial \widetilde{a}_0{}^i}{\partial p_a}\left[\gamma(1/2)\sigma_\kappa^2(\widetilde{a}_0^i)^{\gamma-1}\widetilde{\omega}^2\frac{\vartheta^i}{\overline{\vartheta}}(1-\varphi)^2\right], \end{aligned}$$

which yields:

$$\frac{\partial \widetilde{x}_2^i}{\partial p_a} = \varphi \frac{\partial Y}{\partial \widetilde{\eta}}\frac{\partial \widetilde{\eta}}{\partial p_a} + \frac{\partial \underline{r}}{\partial \widetilde{\eta}}\frac{\partial \widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i + \frac{\partial T}{\partial p_a} + \frac{\partial \widetilde{a}_0{}^i}{\partial p_a}\left[A(1-\varphi) + \underline{r} + \gamma\frac{\vartheta^i}{\overline{\vartheta}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right],$$

where $x_r = \widetilde{\omega}(\phi - A)(1-\varphi) - \frac{1}{2}\sigma_\kappa^2\widetilde{\omega}^2(1-\varphi)^2$.

Concerning the term Δ^c, notice that:

$$\ln\left(1 + \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) exp\left(\alpha^i x_2^i\right)\right) = \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) exp\left(\alpha^i x_2^i\right) + \mathcal{E}(a_0^i),$$

where since $\left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) < 0$ we have $\mathcal{E}(a_0^i) < 0$ an approximation error to the transformation $\ln(1 + x) \approx x$. Using this, we can rewrite:

$$\begin{aligned} \Delta^c &= -\widetilde{\alpha}^i\left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)x_2^i + \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right) exp\left(\alpha^i x_2^i\right) + \mathcal{E}(a_0^i) + \ln(\widetilde{\alpha}^i/\alpha^i) \\ &= \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)\left(exp\left(\alpha^i x_2^i\right) - \widetilde{\alpha}^i x_2^i\right) + \mathcal{E}(a_0^i) + \ln(\widetilde{\alpha}^i/\alpha^i), \end{aligned}$$

using $\ln(\widetilde{\alpha}^i/\alpha^i) \approx -\left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)\frac{\widetilde{\alpha}^i}{\alpha^i}$, we get

$$\Delta^c \approx \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)\left(exp\left(\alpha^i x_2^i\right) - \frac{\widetilde{\alpha}^i}{\alpha^i}(\alpha^i x_2^i + 1)\right) + \mathcal{E}(a_0^i).$$

Using the welfare function, the optimal progressivity p_a solves:

$$\frac{\partial \mathcal{W}}{\partial p_a} = \int s(a_0, \vartheta) \frac{\partial \Delta^{CE}(a_0, \vartheta)}{\partial p_a} \mathcal{G}(a_0, \vartheta) = 0\,,$$

which can be rewritten

$$\int s(a_0,\vartheta)\left[\varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}+\frac{\partial \underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i+\frac{\partial T}{\partial p_a}+\frac{\partial \widetilde{a_0}^i}{\partial p_a}\left[A(1-\varphi)+\underline{r}+\gamma\frac{\vartheta^i}{\overline{\vartheta}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right]+\frac{\partial(\Delta^c/\widetilde{\alpha}^i)}{\partial p_a}\right]\mathcal{G}(a_0,\vartheta)$$
$$= 0\,,$$

or equivalently using $\int s(a_0,\vartheta)\mathcal{G}(a_0,\vartheta) = 1$,

$$\underbrace{\varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}}_{\text{efficiency}}+\underbrace{\frac{\partial \underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\int s(a_0,\vartheta)\widetilde{a}_0\mathcal{G}(a_0,\vartheta)}_{\text{rent extraction}}+\underbrace{\frac{\partial T}{\partial p_a}}_{\text{lump-sum transfers}}$$
$$+\int s(a_0,\vartheta)\left(\underbrace{\frac{\partial \widetilde{a_0}^i}{\partial p_a}\left[A(1-\varphi)+\underline{r}+\gamma\frac{\vartheta^i}{\overline{\vartheta}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right]}_{\text{direct effect on level } a_0^i}+\underbrace{\frac{\partial(\Delta^c/\widetilde{\alpha}^i)}{\partial p_a}}_{\text{direct effect on } \alpha^i}\right)\mathcal{G}(a_0,\vartheta)=0\,.$$

A.1.11 Proof of Lemma 5

Proof. Let us first observed that we can rewrite the utility function as:

$$\max_{k_i^1,b_1^i}\left(\frac{1}{1-1/\sigma}\right)\left[\left(a_0^i-k_1^i-b_1^i\right)^{1-1/\sigma}+\beta\left(\mu_{c_2}^i(k_1^i,b_1^i)-\frac{\alpha_i}{2}\sigma_{c_2}^i(k_1^i,b_1^i)\right)^{1-1/\sigma}\right], \qquad (37)$$

with $k_1^i = \omega_1^i a_1^i$ and $b_1^i = (1-\omega_1^i)a_1^i$. To do that, notice that

$$U\Big(G^{-1}\Big(\mathbb{E}\Big[G\Big(U^{-1}(u_2)\Big)\Big)=\Big(G^{-1}\left(\mathbb{E}\left[G(c_2^i)\right]\right)\Big)^{1-1/\sigma}/(1-1/\sigma)\,,$$

using the fact that $x = -(1/\alpha^i)\ln(1-\alpha^i G(x))$ and $\mathbb{E}\left[G(c_2^i)\right] = (1/\alpha^i)(1-exp(-\alpha^i x_{c_2}^i))$ as shown above, we get:

$$G^{-1}\left(\mathbb{E}\left[G(c_2^i)\right]\right)=-(1/\alpha^i)\ln\left(1-\alpha^i(1/\alpha^i)\left(1-exp(-\alpha^i x_{c_2}^i)\right)\right)=x_{c_2}^i\,.$$

Hence the program of the agent can be rewritten as in (37).

Combining both the first order conditions with respect to k_1^i and b_1^i yields

$$k_1^i=\frac{(\phi-A)(1-\varphi)}{\alpha^i(1-\varphi)^2\sigma_\kappa^2}=\widetilde{\omega}\frac{\vartheta^i}{\overline{\vartheta}}(a_0^i)^{\gamma-1}\,.$$

Using the condition with respect to b_1^i, we get

$$(a_0^i - b_1^i - k_1^i) = x_{c_2}^i (\widetilde{R}\beta)^{-\sigma} ,$$

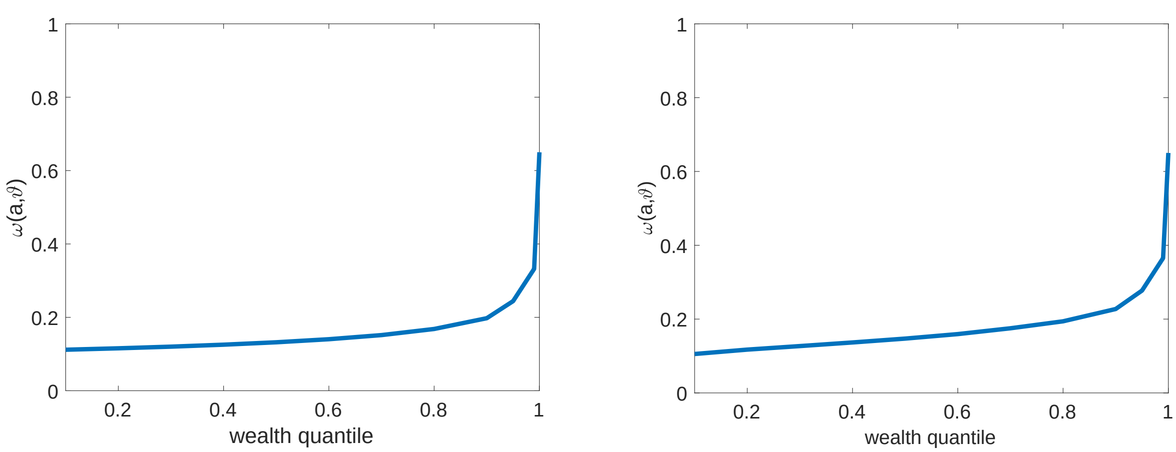
where $\widetilde{R} = \underline{r} + A(1-\varphi)$. Using the expression of $\mu_{c_2}^i = \varphi Y + T + (\underline{r} + A(1-\varphi))b_1^i + (\underline{r} + \phi(1-\varphi))k_1^i$ and $x_{c_2}^i = \mu_{c_2}^i - \frac{\alpha^i}{2}\sigma_{c_2}^i$, we obtain

$$\begin{aligned} b_1^i(\widetilde{R}\beta)^\sigma + \widetilde{R}) &= a_0^i(\widetilde{R}\beta)^\sigma - k_1^i \left[(\widetilde{R}\beta)^\sigma + \underline{r} + \phi(1-\varphi)\right] - \varphi Y - T + \frac{\alpha^i}{2}\sigma_{c_2}^i \\ b_1^i &= (\widetilde{R}\beta)^\sigma + \widetilde{R})^{-1} \left(a_0^i(\widetilde{R}\beta)^\sigma - k_1^i \left[(\widetilde{R}\beta)^\sigma + \underline{r} + \phi(1-\varphi)\right] - \varphi Y - T + \frac{\alpha^i}{2}\sigma_{c_2}^i \right) . \end{aligned}$$

□

A.2 Fit of the Static Model under pure type/scale dependence

Figures 10a and 10b show the fit of the type and scale dependence models regarding the distribution of risky asset shares across the wealth distribution. To fit this shape, we fix inequality to $\eta = 1.4$. The parameter $\gamma = 1.39$ is used to match the average risky asset share of the top 1% in the pure *scale* dependence model. In the pure *type* dependence model, we use the parameter controlling the correlation (to 0.65) between the two distributions and fix the Pareto shape of types to $\varepsilon = 2$. The two models produce an extremely close fit of the observed distribution of the average portfolio allocations across the wealth distribution relative to the one observed in Figure 3.



(a) wealth dependence model

(b) type dependence model

B Empirical Appendix

B.1 Adjusted Survey of Consumer Finance

Throughout the paper, we use the Survey of Consumer Finance (SCF) from 1998 to 2019 (eight waves). Each wave provides cross-sectional data on U.S. households' income and wealth, including detailed information regarding portfolio allocation as well as demographic characteristics.

B.1.1 Details on sampling

Households in the SCF are selected from a double sampling procedure. A first sample is selected from a standard sampling procedure, providing a good representativity of the population. A second sample selects very high income families from the tax records of the Internal Revenue Service (IRS), with some that are also likely to be very wealthy. The SCF weights are used to combine individual characteristics from the two samples to make estimates for the full U.S. population.

B.1.2 Correcting for under-representation and under-reporting

Wealth and income concentration measures from survey data face two common issues: (i) under-representation, meaning that wealth-rich households are generally under-represented in survey data, and (ii) underreporting of assets, meaning that individuals tend to under-report wealth, especially financial wealth.

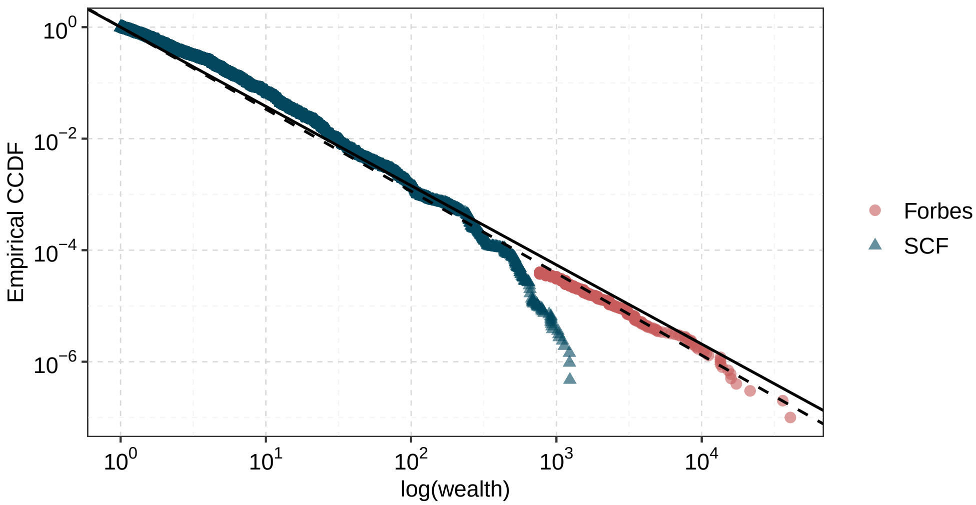
We correct for those issues using the procedure described in Vermeulen (2016). The method is iterative and proceed by assuming that the wealth distribution can be well approximated by a Pareto Law at the top. While this assumption is questionable, most countries admit a linear log-log relationship between wealth level and its empirical CCDF at the top of the distribution, indicating that a Pareto Law can well describe the distribution. The method employs this property to estimate a country-specific Pareto tail and use it to extrapolate estimates for top wealth shares.

First, observation at the top of the wealth distribution in the SCF (above a given threshold of wealth) are supplemented by an external source – such as the Forbes World's Billionaires lists – in order to estimate a Pareto Law using additional observations at the very top. It can be shown that an estimate of the Pareto tail can be obtained by regressing:

$$\ln(n(a_i)/n) = -\eta ln(a_i/a_{min}) \,, \tag{38}$$

where $n(a_i)$ is the number of sample observations that have wealth at or above a_i, i.e. the rank of the observation, and a_{min} is the minimum wealth level at which we assume that the wealth distribution is Pareto (we fix it to 1 million of dollars). $\ln(n(a_i)/n)$ is the log of the relative frequency (or empirical ccdf). Figure 11 shows the resulting Pareto tail $\hat{\eta}$ estimates without and with the 2010 Forbes World's Billionaires lists using the 2010 SCF. From this, we generate corrected for missing value wealth shares by reconstructing a theoretical Pareto distribution at the top.

Figure 11. Pareto shape estimation for the United States using the SCF.



Empirical CCDF in the US combining SCF (2010) and the Forbes World's Billionaires lists. The dashed line corresponds to the estimate of the Pareto tail without the Forbes observations. The solid line includes observations from the rich list.

Second, aggregate estimates from the entire wealth distribution, below and above the threshold of wealth, are computed for total, financial and non-financial assets. Households' wealth are then adjusted such that aggregate estimates from the survey data supplemented with the Forbes's list coincide with national households balance sheet.

The procedure iterates until the distribution of wealth is invariant.[48]

B.2 Measuring Returns to Wealth in the PSID

We use the PSID to compute returns to wealth along the wealth distribution. Our sample spans the period from 1998 to 2019. Every two years, there is a new wave (thus ten in total). The initial period is lost when we compute the returns.

We follow closely the procedure of Fagereng et al. (2020). Our sample considers households whose the head is aged between 20 and 70. This is to ensure that the financial decision maker is the holder of the assets. Moreover, our model abstracts from many decisions that may occur at the end of life (voluntary bequests, health expenditures etc). We restrict our attention to households with at least $1000 of net worth. This ensures that returns to wealth are finite.

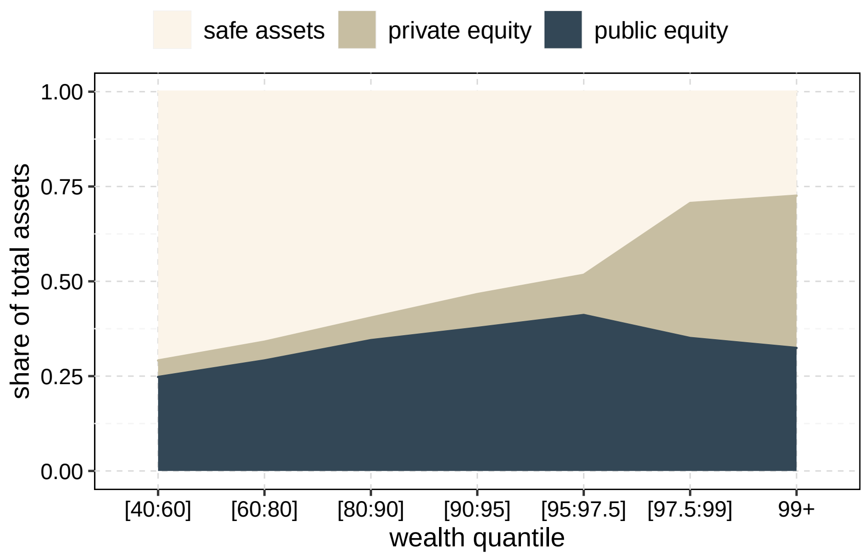
B.2.1 Portfolio composition in the PSID

Before turning to the analysis of returns, we first display in Figure 12 the average portfolio composition across the wealth distribution in the PSID. As can be seen, and consistent with our SCF

[48]While there is no reason to think that this procedure converges to a fixed point, it appears that this is indeed the case for the US economy and all European countries from the HFCS (see the Online Appendix OA 3.1.3).

sample, the share of risky assets is increasing along the wealth distribution, reaching around 60% within the top 5% (25% in private equity and 35% in public equity).

Figure 12. Portfolio composition in the PSID.



B.2.2 Returns to wealth

We compute returns to *net worth* and for three categories: safe assets, public equity and private equity. Following our justification in the core paper, safe assets include housing (primary and secondary residence including rental properties or cottages), riskfree assets (checking/saving accounts, money market funds, certificates of deposits, government bonds, or treasury bills) and other assets (boats, motor homes, cars, cash value in a life insurance policy, a valuable collection, or rights in a trust or estate). Private equity includes businesses and farms, and public equity includes direct holdings in publicly held corporations and indirect holdings in mutual funds, investment trusts and through employer-based pensions or IRAs.

Our definition of net worth is the amount the household would receive if they would sold their assets and paid off all debts associated with the asset. Total liabilities include loans, mortgages, consumer credits and other loans.

We define the total amount of gross assets as:

$$a^{g}_{i,t} = a_{i,riskfree,t} + a_{i,home,t} + a_{i,secondary,t} + a_{i,other,t} + a_{i,priv,t} + a_{i,public,t}\,. \tag{39}$$

The PSID reports inflows and outflows from each assets. Inflows represent all investments, additions and upgrades of assets, while outflows represent all disinvestment, liquidation, and asset sales. We denote $F_{i,l,t}$ the net inflow (total inflows minus total outflows) for asset $l \in \{riskfree, home, secondary, priv, public, other\}$.

Asset values are available for each period for holdings of public equities and for primary residences. Unfortunately, asset values for private businesses and secondary housing are only observed from 2011 onward. Prior to 2011, only the *equity* value (marketable asset value minus total debt associated with the asset $e_{i,l,t} = a_{i,j,t} - d_{j,l,t}$) is observed for secondary housing and private business assets. We impute the asset value using backward induction using $\Delta^a_{i,l,t} = a_{i,l,t} - a_{j,l,t-2} \approx e_{i,l,t} - e_{i,l,t-2}$ and net inflows $F_{i,l,t}$, such that:

$$a_{i,l,t-2} = a_{i,l,t} - \Delta^a_{i,l,t} - F_{i,l,t}\,, \qquad l \in \{priv, secondary\}\,,$$

which implies that any variation in debt $d_{i,l,t}$ translates one-to-one into variations in the value of the asset $a_{i,l,t}$. We now compute the *pre-tax* returns to wealth using our PSID sample. To do so, we make a number of steps to compute capital gains, income and costs.

Capital gains We compute (unrealized and realized) capital gains by comparing the value of each asset at two consecutive waves while taking into account inflows and outflows from this asset. Specifically, we compute capital gains of a household i, period t and asset l as follows.

- For primary residence (*home*), unrealized capital gains are defined as the difference between the current marketable value $a_{i,home,t}$ minus the past one (in the previous wave) $a_{i,home,t-2}$. Realized capital gains are defined as the selling price p^{sell}_{home} less the marketable value in the previous wave, $a_{i,home,t-2}$. To isotate the variations due to capital gains, we take into account inflows and outflows. Inflows are all additions and upgrades, denoted $I_{i,home,t}$. Outflows take into account that the stock of housing depreciate, at a rate of $\delta_h = 2.0\%$ which is the average maintenance and repair cost over the marketable asset value available from 2005 onward, such that: $F_{i,home,t} = I_{i,home,t} - \delta_h(a_{i,home,t} + a_{i,home,t-2})$. We obtain that capital gains for the primary residence are given by:

  $$R^K_{i,home,t} = \frac{\mathbb{1}_{\{sold=1\}} p^{sell}_{home} + \mathbb{1}_{\{sold=0\}} a_{i,home,t} - a_{i,home,t-2} - F_{i,home,t}}{2}\,,$$

  where we divide by 2 to annualize the capital gains.

- For other assets (public equity, private equity and secondary housing), unrealized and realized capital gains $R^K_{i,l,t}$ are defined as the difference between the current marketable asset value and the one observed in the previous wave, minus net inflows $F_{i,l,t}$, such that:

  $$R^K_{i,l,t} = \frac{a_{i,l,t} - a_{i,l,t-2} - F_{i,l,t}}{2} \qquad l \in \{priv, public, secondary\}\,,$$

  Notice that in case of new business acquisition of a value equals to A, we obtain $F_{i,priv,t} = A$ $a_{i,l,t-2} = 0$ and $a_{i,l,t} = A$. Therefore, $R^K_{i,priv,t} = 0$. We obtain the reverse in case of liquidation. As such, our measure of capital gains accommodates for new acquisitions and total sales.

Capital income For each category, capital income refers to the sum of capital income earned by the head and the spouse. We define capital income for each asset category l as follows.

- Capital income from primary residence, $R^I_{i,home,t}$, takes into account maintenance cost and the rental value in the calculation. Lacking evidence on those two components, we assume:

$$R^I_{i,home,t} = r_h a_{i,home,t-2} + inc^{rent}_{i,home,t} - \delta_h a_{i,home,t-2}$$

  where $inc^{rent}_{i,housing,t}$ is the rental income reported to all housing assets. We attribute rental income to the primary residence when the household has no secondary residence, in such case we subtract $0.5\, utils_{i,home,t}$ from the rents, and to secondary residence otherwise. Finally, consistent with Flavin and Yamashita (2002), we assume that the housing yield have an interest component with rate $r_h = 5\%$. Again, δ_h denotes the depreciation rate. Apart from the fact that we directly observe net inflows, it should be noticed that our approach is different from Flavin and Yamashita (2002). Adopting exactly their specification would shift upward returns for primary residence, but turns to be less consistent with values reported in Fagereng et al. (2020) and Bach et al. (2020).

- Capital income from secondary residence, $R^I_{i,secondary,t}$, depends on whether the property is rented, occupied by the household, or not occupied. We assume that

$$R^I_{i,secondary,t} = \mathbb{1}_{\{occupied\}} r_h a_{i,secondary,t-2} + \mathbb{1}_{\{rented\}} inc^{rent}_{i,secondary,t} - \delta_h a_{i,secondary,t-2}\}$$

- Capital income from private equity businesses, $R^I_{i,priv,t}$ is computed as follows. Private equity income is split evenly between labor and asset income if the household actively participates in a private business and only to asset income otherwise.

- Capital income from riskfree assets, $R^I_{i,riskfree,t}$, is obtained from interest income reported in the PSID. As there is no distinction between the fraction of interest income coming from safe assets relative to public equity, we proceed as follows. We assume that interest income from riskfree assets is given by the maximum between the reported interest income $inc^{interest}_{i,t}$ and income derived from the 1-year Treasury bill secondary market rate times the value reported from bond interest, i.e. $r^{treasury}_{i,t} \times a_{i,bond,t}$, such that: $R^I_{i,riskfree,t} = \min\{r^{treasury}_{i,t} \times a_{i,bond,t}, inc^{interest}_{i,t}\}$. A positive difference $\Delta^{interest}_{i,t} = inc^{interest}_{i,t} - R^I_{i,riskfree,t}$ is then associated to public equity. Results are not very sensitive to this assumption.

- Capital income from public equity, $R^I_{i,public,t}$, is equal to the sum of dividends, income from stocks held into IRAs and pension accounts, other interest income and trusts. We obtain income from public equity as: $R^I_{i,public,t} = \Delta^{interest}_{i,t} + inc^{dividend}_{i,t} + inc^{IRA}_{i,t} + inc^{trust}_{i,t} + inc^{other\,fin}_{i,t}$.

Capital debt cost The last important component of our definition of returns are debts. For primary residence, the cost of debt, $R^D_{i,home,t}$ corresponds to the repayment of mortgages. The PSID contains information for two mortgages. We compute the average mortgage interest rate as a weighted average between them and deflate this rate using the CPI index. We follow Flavin and Yamashita (2002) and assume that interest payment are deductible, such that the household pays a real after-tax interest rate of $r^D_{i,home,t} = \frac{1+(1-\tau)r^{mortgage}_{i,t}}{1+inflation_t} - 1$, where we set $\tau = 33\%$. All other costs of debt are computed assuming an interest rate of 5% on other remaining debts.

Return measure Following Fagereng et al. (2020), our reference measure of return of an asset l is:

$$r_{i,l,t} = \frac{R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t}}{a_{i,l,t-1} + F_{i,l,t}/2} \tag{40}$$

The numerator is the sum of income, $R^I_{i,l,t}$, capital gains, $R^K_{i,l,t}$, minus the cost of debt, $R^D_{i,l,t}$ accrued by household i on asset l in year t. The denominator is defined as the sum of beginning-of-period stock of gross wealth and net flows of gross wealth during the year. In the PSID, wealth is observed only at the time of the interview while income for each asset are observed for the past year. Because the periodicity of the data is biennial, we need to impute the beginning-of-period asset level corresponding to the income derived from the asset. We do so by assuming that beginning-of-period asset is the interpolation between the current wealth level and the wealth level reported in the previous wave, such that: $a_{i,l,t-1} = (a_{i,l,t-2} + a_{i,l,t})/2$. The second term on the denominator, $F_{i,j,t}$, accounts for the fact that asset yields are generated not only by beginning-of-period wealth but also by additions/subtractions of assets during the year. Without this adjustment, we may bias our estimates if the beginning-of-period wealth is small but capital income is large due to positive net asset flows occurring during the period (for example, a business acquisition). As the flows occur during the year, we make the assumption that they occur on average in mid-year.

In equation (40), we express the dollar yield on net worth as a share of gross wealth (or total assets) to ensure that the sign of the return reflect the sign of the yield.

Notice that for *net worth*, we define:

$$r^{networth}_{i,t} = \frac{\sum_l (R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t})}{a^g_{i,t-1} + \sum_l F_{i,l,t}/2} \qquad r^{gross}_{i,t} = \frac{\sum_l (R^K_{i,l,t} + R^I_{i,l,t})}{a^g_{i,t-1} + \sum_l F_{i,l,t}/2} \tag{41}$$

We convert all nominal returns to real returns using the consumer price index (CPI) from the Federal Reserve, using: $\tilde{r}_{i,l,t} = \frac{1+r_{i,l,t}}{1+inflation_t} - 1$.

Trimming Finally, we trim the distribution of returns in each year t and for each asset category l at the top and the bottom by 0.5%. This ensures that there is no outlier polluting the estimates of the mean of returns and aim to reduce measurement errors.

B.2.3 Scale dependence in returns

In this section, we use the PSID to evaluate the presence of scale dependence in the returns to wealth. To do so, we follow Gabaix et al. (2016) and Fagereng et al. (2020) representation and estimate the following statistical model:

$$r_{i,t}^{gross} = \theta^g P_a(a_{i,t-1}^g) + f_t + f_i + \epsilon_{i,t}^g\,, \qquad r_{i,t}^{networth} = \theta^n P_a(a_{i,t-1}^n) + f_t + f_i + \epsilon_{i,t}\,, \tag{42}$$

where $r_{i,t}^{gross}$ and $r_{i,t}^{networth}$ are respectively the return to gross and net wealth, $P_a(\cdot)$ is the percentile of beginning-of-period gross/net wealth (capturing scale dependence), f_i^g, f_i^n are the individual fixed effect (capturing persistent heterogeneity), f_t^g, f_t^n are time fixed effects capturing aggregate return components, and $\epsilon_{i,t}^g, \epsilon_{i,t}^n$ are error terms. *Scale* dependence is measured by the parameter θ^g for gross returns and by the parameter θ^n for net returns, while *type* dependence is captured by the individual fixed effect. Thus, similar to Fagereng et al. (2020), the *scale* dependence (comprising all sources of scale dependence, direct and indirect) parameters are identified from household-specific time variations in the wealth percentile.

Table 10. Scale dependence regression

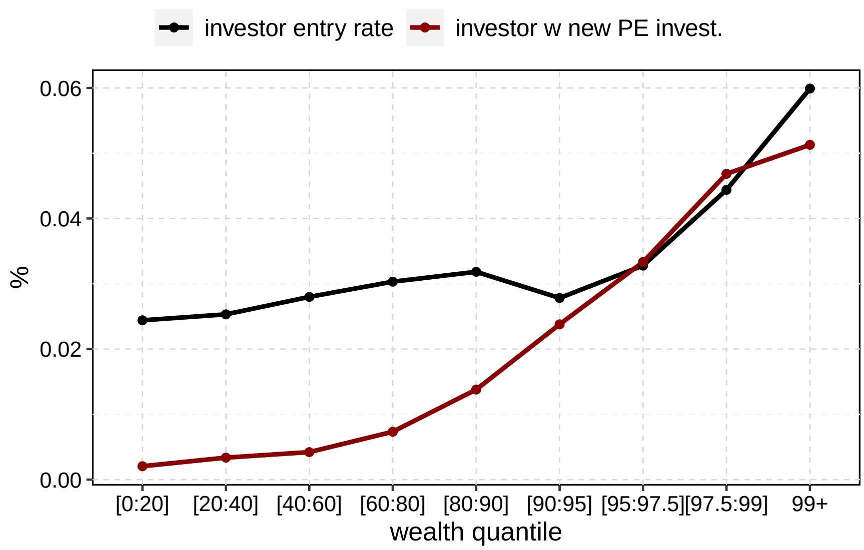
	percentile		log specification	
	Gross Return	Net Return	Gross Return	Net Return
θ	0.152	0.121	0.034	0.024
Time FE	Yes	Yes	Yes	Yes
Observations	9286	7491	9286	7491

We find a positive and statistically significant degree of *scale* dependence in returns to wealth in the PSID, confirming the findings of Fagereng et al. (2020) and Bach et al. (2020) who use Scandinavian administrative data. Moreover, the estimates are surprisingly close to the ones obtained by Fagereng et al. (2020) using Norwegian data. Notice that we obtain a similar qualitative message regarding the relationship between returns to wealth and wealth if we regress $log(a_{i,t-1}^g)$ and $log(a_{i,t-1}^n)$ on gross and net returns, respectively.

B.3 PSID participation and diversification

In Figure 13, we use our PSID sample to show the unconditional average probability to switch from a non private equity investor state to a private equity investor state (black line), and the unconditional fraction of private equity investors who invest in a new additional private equity investment (red line). Focusing on the black line, it is apparent that there is an increasing and convex relationship between risky investment participation and wealth, a feature which is consistent with Hurst and Lusardi (2004). Focusing on the red line, it appears that, conditional on being already private equity investor, new additional investments in private equity occur essentially at the very top. We interpret this as evidence for diversification among the wealthy households, which is consistent with the recent paper by Penciakova (2018). We now investigate these relationships while controlling for household's characteristics.

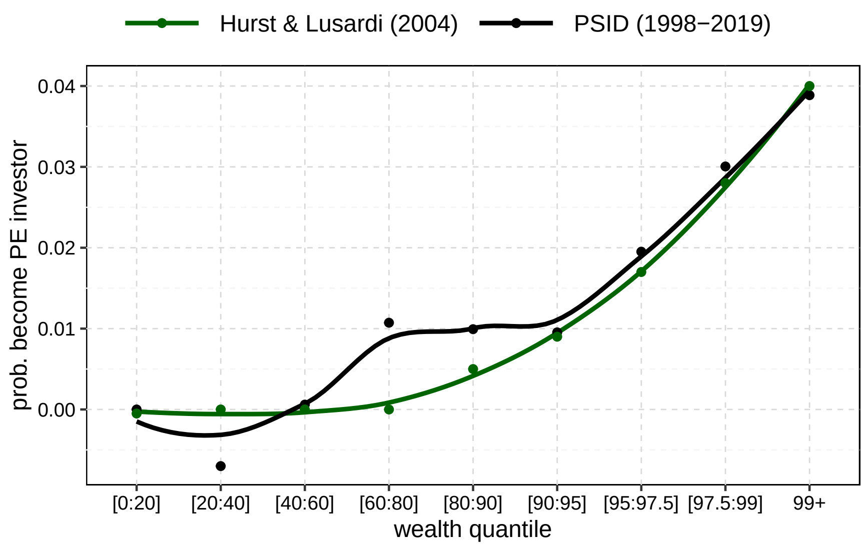
Figure 13. Private equity (PE) investment participation and diversification in the PSID.



Conditional relationships In the main text, we model the fact that the entry into "investor" state increases with wealth. To this, Hurst and Lusardi (2004) estimate the probability to become a private equity investor as a function of wealth while controlling for household's characteristics. They then compute the predicted probability to participate into private equity investments along the wealth distribution. They find that only the wealthy households (above the top 95^{th} percentile) are more likely to switch to private equity business ownership when their wealth increases. Following their lead and using similar controls, we update their estimation using our sample period. In Figure 14, we report the predicted probability as estimated in Hurst and Lusardi (2004) as well as our update. Results are found to be very similar.

Finally, we report in Figure 15 the predicted probability of acquiring a new private equity business investment, conditional on being an investor. Consistent with what is observed in the Survey of Consumer Finance, we find that the probability to invest in multiple businesses increases at the top of the wealth distribution. While we cannot control for household's characteristics in the SCF, our results from the PSID reveal that this relationship between new acquisition of a private equity business investment and wealth is robust and do not hinge on an observed household characteristic (which may be a proxy for "type") only.

Figure 14. *Scale* dependence in investment participation.



Predicted probability relative to the mean.

C Quantitative Appendix

In this section, we show additional details regarding the quantitative model, the calibration, the computational algorithm and additional moments of interest.

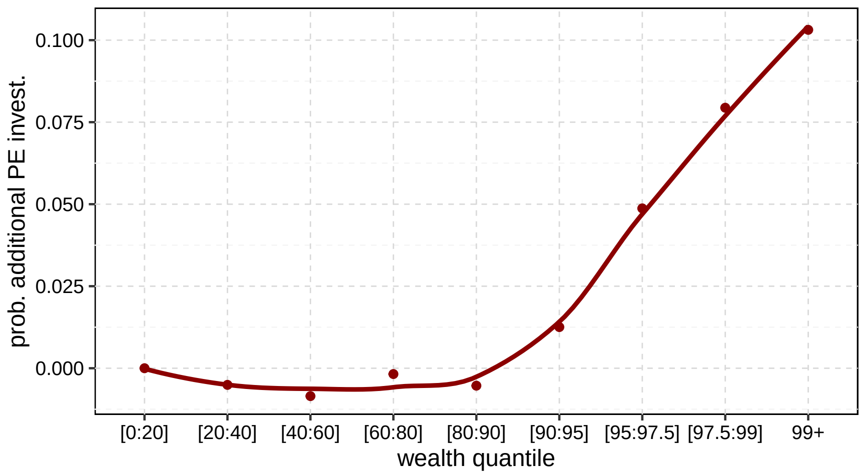
C.1 Calibration

Benchmark economy (M1) Table 11 displays the scale dependence in the intensive margin of portfolio allocation, conditional on being an investor, in the data and in the benchmark economy. In the benchmark economy, the parameter $\{\underline{a}_\omega, \omega_1, \omega_2, \gamma_\omega\}$ are used to match the following moments. $\underline{a}_\omega$ is chosen to correspond to the wealth level of the 70^{th} wealth percentile in the model, such that there is no scale dependence in risky portfolio observed below this percentile. The level and the shape parameters are endogenously set to $\omega_1 = 0.072$ and $\gamma_\omega = 0.30$ in order to replicate the average share invested in risky equity (through additional investments) of 11% for households within the [95-97.5] wealth quantile, and of 20% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.20$ is set to guarantee that the wealth distribution is stationary and to match the average risky portfolio share above the top 0.1%, as observed in Figure 5.

Table 11. Resulting scale dependence in risky portfolio share: benchmark model and data

Wealth quantile	[0–80]	[90–95]	[95–97.5]	[97.5–99]	[99–99.9]	top 0.1%
SCF average risky share	0%	6%	11%	15%	20%	20%
Benchmark model $\omega(a,1)$	0%	6.7%	10.3%	12.6%	16.9%	20%

Figure 15. *Scale* dependence private equity business investment.



Predicted probability relative to the mean.

Table 12 reports the scale dependence in the risky investment participation in the benchmark economy and in the PSID (see Figure 14 and Hurst and Lusardi (2004)). In the benchmark economy, the parameters $\{\underline{a}_{\lambda}, \lambda_1, \lambda_2, \gamma_{\lambda}\}$ are used to match the following moments. $\underline{a}_{\lambda}$ is chosen such that it corresponds to the wealth level of the 80^{th} wealth percentile in the model, such that there is no scale dependence in the risky investment participation below this percentile. In this case, the participation rate of 1.8% is only generated through the process governing ϑ-types. The level and the shape parameters of the transition probability with respect to wealth are endogenously chosen to be $\lambda_1 = 0.071$ and $\gamma_{\lambda} = 0.30$ to replicate the average transition rate of 3.2% for households within the [95-97.5] wealth quantile, and of 6.1% for households within the [99-99.9] wealth quantile. A maximal value of $\lambda_2 = 0.045$ is set to guarantee that the wealth distribution is stationary and to match the maximum transition rate of 7% at the very top (within the top 0.1%).

Table 12. Resulting scale dependence in the risky investment participation: benchmark model and data

Wealth quantile	[0–80]	[90–95]	[95–97.5]	[97.5–99]	[99–99.9]	top 0.1%
PSID entry rate into "investor" state	1.8%	2.1%	3.2%	4.5%	6.1%	7.0%
Benchmark model entry rate $\underline{\pi_{\vartheta}} + \lambda(a)$	1.8%	2.2%	3.2%	4.4%	6.3%	7.0%

Scale model (M2) In this model, we follow Hubmer et al. (2020) and match the increasing average risky asset shares along the wealth distribution through scale dependence only. In this model, the only source of scale effects comes from the intensive margin of risky portfolio shares such that equation (43) in the main text becomes:

$$\omega(a, \vartheta) \equiv \omega(a) = \underline{\omega} + \varpi(a) = \underline{\omega} + \min\left\{\omega_1 \left(\max\left\{a - \underline{a}_{\omega}, 0\right\}\right)^{\gamma_{\omega}}, \omega_2\right\}, \tag{43}$$

The parameters $\{\underline{\omega}, \underline{a}_\omega, \omega_1, \omega_2, \gamma_\omega\}$ are recalibrated to replicate the following moments. $\underline{a}_\omega$ is chosen to correspond to the wealth level of the 80^{th} wealth percentile in the model, such that there is no scale dependence in the risky portfolio share below this percentile. In this case, the average risky portfolio share is fixed to $\underline{\omega} = 10\%$, following our estimates from the SCF. The level and the shape parameters of the portfolio allocation with respect to wealth are endogenously set to $\omega_1 = 0.057$ and $\gamma_\omega = 0.57$ to replicate an average portfolio share of 37% for households within the [95-97.5] wealth quantile, and of 60% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.75$ is set to guarantee that the wealth distribution is stationary and to match the maximum average portfolio share invested at the very top (within the top 0.1%).

Type model (M3) In the type model, there are no scale effects ($\lambda(a) = \varpi(a) = 0$). In this model, we recalibrate $\underline{\omega}$ such that the average risky share in the economy corresponds to the one observed in the SCF. The switching probability $\underline{\pi}_\vartheta$ is set to match the fraction of investors.

C.2 Computational appendix

C.2.1 State space and grid definition

In our model, an household is fully characterized by a state vector $\mathbf{s} = (a, \vartheta, h, j) \in \mathbb{S} \equiv \mathbb{R}^+ \times \Theta \times \mathbb{H} \times \mathcal{J}$. We compute the household problem using a grid of asset $\mathbf{a}$ of 350 points (adding more points only marginally increase our accuracy) spaced according to an exponential rule. We truncate the grid for asset to $A_{\max} = 1500000$ and we impose the borrowing constraint such that $A_{min} = 0$. Due to the finite upper bound on wealth and the Pareto property coming from heterogeneous returns (see Benhabib et al. (2011)), the resulting distribution of wealth is not always ergodic. In the main text, we describe a procedure to verify the size of the approximation error generated by this upper bound on wealth and conclude that it is small.

We discretize the process h with 9 grid points spaced according to the following quantiles of the persistent income component distribution: $\mathbf{q}_h = [0.01 \;\; 0.15 \;\; 0.30 \;\; 0.45 \;\; 0.60 \;\; 0.75 \;\; 0.9 \;\; 0.95 \;\; 0.99]$. Following Hubmer et al. (2020), the values h corresponding to these quantiles follow the log-normal/Pareto mixture described in the core paper. Given the calibration, h values attached to households within the top 10% highest earners are drawn from a Pareto distribution, and from a log-normal distribution otherwise.

C.2.2 Algorithm

We organize the algorithm in steps.

1. Initialize a full dimension grid space over asset values (a), productivity level (h), age bracket (j), and innate investment skill (ϑ).

2. Guess initial tax rates $\boldsymbol{\tau}$ and equilibrium quantities $\{\frac{X}{L}, \underline{r}\}$. Compute p and w.

3. Given prices, solve the consumption-saving-leisure problem. We use a modified version of the EGM algorithm introduced by Carroll (2006).

   Specifically, the budget constraint can be written as $c = \frac{why\zeta_j(1-\tau_w)\ell - a' + \mathcal{F}(a,\vartheta)}{1+\tau_c}$. Given the utility function $u(c,\ell)$, the optimality conditions are given by

   $$\ell(c) = c^{-\sigma\lambda}\left[\frac{why\zeta_j(1-\tau_w)}{\chi(1+\tau_c)}\right]^{\lambda}, \qquad c^{-\sigma} = \beta(1-d_j)(1+\tau_c)\mathbb{E}[v_a(a',\vartheta',h)]\,.$$

   We compute v_a numerically. To use the endogenous grid method, we invert the consumption-leisure intratemporal condition and express ℓ as a function of c. We plug the solution in the budget constraint, such that:

   $$(1+\tau_c)c + a' = [why\zeta_j(1-\tau_w)]^{1+\lambda}\left(\frac{1}{(1+\tau_c)\chi}\right)c^{-\sigma\lambda} + \mathcal{F}(a,\vartheta)\,.$$

   To gain in speed and avoid any root-finding within the policy function iteration, we pre-compute all possible realizations of (c,ℓ) on an exogenous grid of cash on hand $\mathcal{F}(a,\vartheta)$. The EGM is performed on an endogenous grid defined as:

   $$\widetilde{\mathcal{F}}(a,\vartheta) = (1+\tau_c)c(a',\vartheta,h) + a' - [why\zeta_j(1-\tau_w)]^{1+\lambda}\left(\frac{1}{(1+\tau_c)\chi}\right)c(a',\vartheta,h)^{-\sigma\lambda}\,.$$

4. Construct the transition matrix $\mathbf{M}$ generated by Π_h, Π_ϑ, Π_κ and Π_y, $a'(\mathbf{s})$ and $\ell(\mathbf{s},y)$. Compute the associated stationary measure of individuals $\mathcal{G}(\mathbf{s})$; by first guessing an initial distribution, and then by iterating on $\mathcal{G}'(\mathbf{s}) = \mathbf{M}\mathcal{G}(\mathbf{s})$ until convergence.

5. Compute the resulting total efficiency units of capital X, total labor supplied L, the return component $\underline{r}$ and government expenditures and revenues.

6. With a relaxation, update the vector of prices; $\{p,w\}$ are obtained using the first order conditions of the representative final good producer, $\underline{r}$ is adjusted to ensure that total returns to capital distributed in the economy is equal to total product of capital in the economy, and the tax rate τ_w is adjusted to balance the government budget if necessary.

   Back to step 2 and iterate until convergence on the equilibrium prices is reached.

C.3 Wealth Mobility

In this section, we investigate how the intra-generational wealth mobility matrix in our model alternatives compares with its empirical counterpart. We take as a reference the estimates by Klevmarken et al. (2003), who compute a five-state (quintiles) five-year transition matrix from the 1994–1999 PSID waves. Table 13 reports the results for the benchmark economy (M1), the type-model (M2) and the scale-model (M3). We find that the three models overstate the persistence of wealth-rank in the top quintiles while being broadly consistent with the U-shaped diagonal transition rate. Interestingly, adding portfolio heterogeneity helps in generating an empirically consistent wealth mobility. Overall, we find that it is hard to distinguish models M1, M2 and M3 based on the resulting wealth mobility matrix.

Table 13. Wealth mobility: data and model

5-years transition	Diagonal element (quintile – quintile)				
	Q1 – Q1	Q2 – Q2	Q3 – Q3	Q4 – Q4	Q5 – Q5
PSID (1994-1999), Klevmarken et al. (2003)	0.58	0.44	0.42	0.48	0.71
M1 – *Benchmark* model	0.57	0.46	0.34	0.42	0.79
M2 – *Scale*-model	0.57	0.37	0.34	0.47	0.81
M3 – *Type*-model	0.57	0.45	0.41	0.42	0.78
M5 – No portfolio heterogeneity	0.64	0.40	0.52	0.60	0.84