Redistribution of Return Inequality*

Karl Schulz†

University of Mannheim

March 16, 2021

Abstract

Wealthier households obtain higher returns on their investments than poorer ones. How should the tax system account for this return inequality? I study capital taxation in an economy in which return rates endogenously correlate with wealth. The leading example is a financial market, where the rich acquire more financial information than the poor. Contrary to conventional wisdom, rather than calling for more redistribution, the presence of this scale dependence provides a rationale for lower marginal tax rates. The endogeneity of returns generates an inequality multiplier effect between wealth and its returns. Therefore, standard elasticity measures that determine the responsiveness of capital to taxes must be revised upwards. At an aggregate level, a rise in redistribution induces a compression effect on the distribution of pre-tax returns. In the financial market, I identify general equilibrium trickle-up externalities that provide a force for more redistribution relative to the partial equilibrium. Finally, I estimate partial and general equilibrium responses and demonstrate the quantitative importance of scale dependence for tax policy.

Keywords: Optimal Taxation, Capital Taxation, Heterogeneous Returns, Wealth Inequality, General Equilibrium, Asset Pricing, Private Information, Financial Literacy

JEL Classification: H21, H23, H24, D31, G11, G12, G14, G53

*I thank Eduardo Dávila, Hans Peter Grüner, Eckhard Janeba, Thomas Piketty, Morten O. Ravn, José Víctor Ríos Rull, Dominik Sachs, Arthur Seibold, Sebastian Siegloch, Konrad Stahl, Holger Stichnoth, Alisa Tazhitdinova, Michèle Tertilt, Ernst-Ludwig von Thadden, Nicolas Werquin, and numerous seminar and conference participants for helpful comments and discussions. I gratefully acknowledge the support by the University of Mannheim's Graduate School of Economic and Social Sciences, funded by the German Research Foundation (DFG).

†University of Mannheim, Center for Doctoral Studies in Economics, L7 3-5, 68131 Mannheim, Germany, karl.schulz@gess.uni-mannheim.de.

1 Introduction

Over the last decades, numerous countries have seen a rapid rise in wealth inequality. In the US, for example, the wealth share of the top 0.1% has tripled over the past forty years (Saez and Zucman (2016)). Heterogeneity and persistence in the idiosyncratic returns to wealth have been successful in explaining the observed thick tail in the wealth distribution. Such "type dependence" can, for instance, plausibly arise from differences in entrepreneurial ability. To account for the cross-sectional dynamics in inequality, one needs to add to standard random growth models a positive correlation between income and its return (Gabaix, Lasry, Lions, and Moll (2016)). This "scale dependence" may arise from various sources. Most prominently, Piketty (2014) argues that wealthier households obtain higher rates of return than poorer ones both across and within asset classes because they can take more risks and hire skilled financial advisers. A recent wave of empirical papers documents the prevalence of this scale dependence.[1]

A well-known result in public finance is that exogenous inequality in capital gains (type dependence) justifies the positive taxation of capital. However, little is known about the policy implications of scale dependence despite its potential to considerably amplify wealth inequality. How *should* a government design the tax system in the presence of scale dependence to reduce inequality? When the rich become richer because they are rich, *should* they pay confiscatory taxes? Which sources of inequality *should* governments address, and which not? *Can* the government alter the inequality of pre-tax return rates? To answer these questions, I introduce endogenously formed return inequality into the optimal taxation of capital. As a leading example, I follow the argument by Piketty (2014), as originated by Arrow (1987): Wealthy households hold a sizable portfolio on financial markets. Therefore, they have a high incentive to purchase information about the stochastic fundamentals, which drive stocks' payoffs. As a result of their better knowledge, they make more informed investment decisions (portfolio choices) and obtain higher rates of returns on their financial investments than poorer households.[2]

According to conventional wisdom, one might expect that, when the rich endogenously obtain higher rates of return than the poor, this additional source of inequality provides a rationale for higher tax rates. I show how to characterize the optimal tax system with and without scale

[1] For instance, see Bach, Calvet, and Sodini (2020) and Fagereng, Guiso, Malacrino, and Pistaferri (2020).

[2] I consider knowledge acquisition in the financial market as the leading example. It generates qualitatively the same endogenous return inequality as other potential channels would do, e.g., stock market participation costs, housing, liquidity constraints, and insurance against consumption risk. Moreover, it fits well into the empirical setting I consider later. The positive and normative implications for capital taxation remain the same irrespective of the underlying mechanism that generates scale dependence.

dependence in terms of empirically observable sufficient statistics. Firstly, the optimal capital tax is inversely related to the elasticity of capital. Secondly, the optimal tax is increasing in observed inequality. Conditional on the observed inequality and elasticities, the optimal tax formulas are the same irrespective of whether capital gains are exogenous (no scale dependence) or endogenously formed (scale dependence).

These two measures that determine the optimal capital tax depend, however, on the process under which returns form as well as on the underlying tax code. Scale dependence may raise observed wealth inequality, which calls for higher capital taxes relative to a setting with exogenous capital returns. In turn, the rise in capital taxes reduces inequality. Simultaneously, the wealth elasticity has to be revised upward in the presence of endogenous capital gains, as I describe later. This provides a force for lower marginal tax rates.

I show that, under certain conditions, the second force dominates the first one for a given observed inequality. Perhaps surprisingly, the optimal marginal tax rate on capital is, therefore, lower with scale dependence than without. In other words, when the rich become richer, not due to their exceptional talent but simply because they are affluent, the government should redistribute less. For a given wage distribution, I find a full neutrality result: Conditional on the primitives of the economy, the introduction of scale dependence does not alter the optimal tax rate. The rise in inequality just offsets the increase in the capital elasticity. Altogether, depending on the comparative statics exercise, scale dependence is either neutral or provides a rationale for lower taxes.

This conclusion is at odds with Piketty (2014) (Chapter 12), who uses scale dependence as an argument for more redistribution (via a progressive wealth tax). It does not mean that return inequality per se leads to lower capital taxes relative to an economy without return inequality. It instead provides a note of caution regarding the policy implications of different *sources* of return inequality. Also note that the results simply arise from pure efficiency considerations. From an equity perspective, relative to type dependence, the optimal tax rate with scale-dependent capital gains might be even lower. The society might put a higher marginal welfare weight on households who took the effort to increase their rates of return (e.g., via the acquisition of financial knowledge). Hence, they may deserve these rates of return, as these reflect a fair reward for effort, compared to households who were just lucky enough to be talented investors (type dependence).

Under scale dependence, there is a two-way interplay between taxes and capital gains. On the one hand, capital gains and their distribution across households shape the wealth distribution, which serves as a critical primitive for designing a tax system (first channel). This channel is also present in standard taxation models with exogenous return inequality (type dependence). On the other hand, if the savings elasticity is non-zero, taxes will affect the incentives to save. In the presence of scale dependence, however, higher savings induce an increase in pre-tax return rates which yields a convex relationship between pre-tax capital gains and savings (second channel).

According to the first channel, taxes are a function of capital return inequality. By the second channel, the distribution of pre-tax capital returns is endogenous to the tax system. I demonstrate that tax reforms induce a compression effect on the pre-tax return distribution. A rise in redistribution reduces the variance of pre-tax returns (equity gains). However, this compression comes along with the cost of lowering mean pre-tax returns (efficiency costs). Consequently, scale dependence gives rise to a novel model-inherent trade-off for tax policy regarding pre-tax return rates. This distributional effect of tax reforms may also provide a source for empirically identifying the magnitude of scale dependence. If there is only type dependence and no scale dependence, the mean pre-tax returns and their variance should not respond to tax reforms.

In the leading example of the financial market, the size of financial portfolios change with capital tax rates. As a reaction, the amount of acquired information, e.g., via financial advisory or financial education, and the optimally chosen portfolio composition adjust. This leads to the described adjustment in the distribution of pre-tax return rates. To the best of my knowledge, this is the first paper that addresses this simultaneous link between redistribution and financial market outcomes.

In the presence of endogenous pre-tax return rates, the standard income and substitution effects from tax reforms have to be augmented by inequality multiplier effects. To provide an example, suppose that the government decreases the capital tax of an individual. Assuming that the substitution dominates the income effect, she saves more. However, when the amount of investment and its return endogenously correlate, the latter also rises (in the leading example, because the individual increases her financial knowledge). Now, she earns more on every dollar she invests in future consumption. In other words, saving money pays off to a greater extend. Therefore, the individual saves more, which in response increases her returns and so on. The responsiveness of the own returns (own-return elasticity) measures this inequality multiplier effect. This observation implies a Le Chatelier principle for capital (see Samuelson (1948)). Due to the endogeneity of pre-tax return rates, capital responds in the long-run more elastic than in the short run for fixed return rates. As a result, one will underestimate capital elasticities (for instance, using short-run data) if one does not account for the (long-run) adjustment in pre-tax return rates.

These observations hold under the partial equilibrium assumption of a small open economy and in general equilibrium. In general equilibrium, tax reforms also affect aggregate variables that feed back into the return functions of households. In the financial market example, the equilibrium stock price is an aggregation of information and risk-taking, which both correlate with aggregate wealth. Thus, a household's return on investment is not only a function of her own savings but also of others'. Then, asides from an inequality multiplier effect, a tax reform also induces cross-return effects. The reasoning is as follows. A tax reform changes the respective household's savings and returns (due to the altered financial knowledge). As her savings adjust, in general equilibrium, the returns of others and, hence, their savings change as well. In response, this feeds back into the return of the first household. I measure these general equilibrium externalities in terms of novel cross-return elasticities and identify trickle-up forces that call for higher taxes in general than in partial equilibrium.

To quantify the importance of the scale dependence, I estimate own- and cross-effects between returns and endowment sizes from panel data on the returns of US private foundations. This idea is similar to Piketty (2014) who descriptively documents the amount of scale dependence using return data from US universities. Although universities and foundations are institutional investors who potentially behave differently on the financial market, they may serve as a reasonable proxy for wealthy investors.

Since the foundations' endowment size is considerably larger relative to household wealth, the amount of scale dependence is likely to be underestimated in the data. I find a statistically significant point estimate for the lifetime own-return elasticity of 0.1. This estimate is substantially lower than the one I retrieve from the study by Fagereng et al. (2020) (0.9). The adjustment of capital elasticities and optimal tax rates resulting from this conservative amount of endogeneity in capital gains is, nonetheless, economically sizable. Relative to a setting without scale dependence, the revenue-maximizing linear capital gains tax declines by six percent. For a higher own-return elasticity of 0.5, which is between the estimate obtained here and the one in Fagereng et al. (2020), the optimal capital gains tax decreases by more than 25% (17 percentage points).

I find statistically significant but economically small cross effects. This suggests no or negligible general equilibrium forces. The point estimates support some features of the financial market. Negative cross-effects from the top of the wealth distribution indicate the presence of trickle-up externalities. Moreover, there are slightly insignificant decreasing returns to scale in the own-return elasticity.

Finally, I calibrate the statistical model that empirically describes the return rate formation to match empirical moments in the Survey of Consumer Finances (2016). The model performs surprisingly well in explaining the degree of return inequality in the US. Using these data, I simulate the aggregate responses of pre-tax return rates to various reforms of the current US capital gains tax scheme. I find large revenue gains from raising the capital gains tax. However, these gains decline in the amount of scale dependence, reflecting the efficiency costs resulting from the adjustment of pre-tax return rates. This decline is quantitatively considerable.

The incidence on the level and the dispersion of pre-tax return rates depends on the type and magnitude of tax reforms. Rises in the taxation of the rich evoke quantitatively larger responses than tax cuts for the poor. The level of pre-tax return rates is more affected than their dispersion. This observation indicates that the impact of scale dependence on the tax policy's efficiency margin is more substantial than the equity margin.

Related literature. This paper relates to four strands of the literature. Firstly, I add to the sizable literature on capital taxation. As shown by Saez (2002), return inequality provides an essential justification for why capital taxes should not be zero, unlike in Atkinson and Stiglitz (1976), Chamley (1986), and Judd (1985). So far, the focus in the literature has been on return inequality that arises from type dependence. For instance, Shourideh (2012), Saez and Stantcheva (2018), and Guvenen, Kambourov, Kuruscu, Ocampo-Diaz, and Chen (2019) allow return rates to exogenously differ across households and study the equity and efficiency implications of capital taxation. Gerritsen, Jacobs, Rusu, and Spiritus (2019) analyze capital taxation under type and scale dependence separately. They solely aim to show that capital taxes should be non-zero in both settings without studying the differential implications of type and scale dependence.

I provide a comprehensive analysis of capital taxation in the presence of scale dependence and place a particular focus on the positive and normative implications (relative to a setting without scale dependence). Opposed to Gerritsen et al. (2019), I also microfound this scale dependence. Moreover, I introduce scale dependence into two well-known frameworks: the dynastic framework of linear wealth taxation by Piketty and Saez (2013) and the canonical Mirrlees (1971) model of nonlinear capital income taxation, as in Farhi and Werning (2010). Using the perturbation techniques introduced in Piketty (1997), Saez (2001), and, more recently, Golosov, Tsyvinski, and Werquin (2014), I characterize the optimal linear and nonlinear capital taxation. Besides, I allow for uncertainty (e.g., Aiyagari (1994)) and full intergenerational dynamics by restricting attention to simple tax instruments. Similarly, I separate the nonlinear taxation of labor and capital income. These restrictions allow me to derive a clear-cut characterization of the respective tax systems. However, the main conclusions regarding the presence of scale dependence should carry over to a fully optimal mechanism as considered in the new dynamic public finance literature (see Golosov, Tsyvinski, Werning, Diamond, and Judd (2006) for a review).

Secondly, my paper links to the literature on redistributive taxation in general equilibrium. Rothschild and Scheuer (2013), Ales, Kurnaz, and Sleet (2015), and Sachs, Tsyvinski, and Werquin (2020) extend the original framework by Stiglitz (1982). Deploying the techniques in Sachs et al. (2020), I am, to the best of my knowledge, the first one to provide a thorough analysis of the nonlinear capital tax incidence and optimal capital taxation in general equilibrium. Thereby, I extend the well-known concepts of own- and cross-wage elasticities that matter for labor income taxation to pre-tax return rates in the context of capital taxation.

Thirdly, I add to the literature on financial knowledge in partial (e.g., Arrow (1987) and Lusardi, Michaud, and Mitchell (2017)) and general equilibrium (e.g., Grossman and Stiglitz (1980), Verrecchia (1982), Peress (2004), Kacperczyk, Nosal, and Stevens (2019)). I am not aware of another paper formalizing a link between redistribution and informational efficiency in Grossman and Stiglitz (1980) financial markets. The idea that capital taxes affect the accumulation of financial knowledge is, however, similar to the literature on taxation and human capital (for recent examples, see Krueger and Ludwig (2013), Findeisen and Sachs (2016), and Stantcheva (2017)). Also, notice that the implications of scale dependence for capital taxation derived in this paper are similar to those of superstar compensation schemes for labor income taxation (see Scheuer and Werning (2017)). Whereas superstar effects mostly manifest at the top of the income distribution, the empirical evidence presented here and in earlier studies suggests that scale dependence is widely disseminated throughout the wealth distribution.

Fourthly, in my empirical analysis of a large panel of US foundations, I document the prevalence of scale dependence and, more generally, return inequality as in Yitzhaki (1987). More recently, Bach et al. (2020) and Fagereng et al. (2020) document scale dependence with Scandinavian data. By providing estimates of own- and cross-return elasticity, I also add to the empirical literature on the estimation of capital income and elasticities. I survey this literature in Section 3.1.

Outline of the paper. The paper is structured as follows. First, I establish the main findings in a simple, conceptual framework (Section 2). In Section 3, I describe the empirical implications and propose a direct estimation of own- and cross-return effects with returns data from US private foundations. In Section 4, I calibrate the statistical model from Section 3 to the US economy. Then, I analyze the aggregate incidence of reforming the current US tax code. Section 5 concludes. I relegate all relevant proofs, model extensions, and the microfoundation to the Appendix.

2 The Model

2.1 A Conceptual Framework

In this section, I describe a simple framework to think about capital taxation under the presence of scale dependence. Suppose there is a unit measure of households $i \in [0,1]$ that differ in their labor earnings ability w_i. Aside from working l_i hours (in period 1), household i saves a_i (for period 2). In line with the microfoundation below, a household can increase the return on savings by taking effort x_i at some cost.[3] Under standard monotonicity conditions ($\frac{dx_i}{dw_i} > 0$), this setting gives rise to *scale dependence*. That is, the household's rate of return on savings, $r_i(a_i)$, is increasing in the amount of savings ($\frac{dr_i}{da_i} > 0$). Observe that this does not rule out the presence of type dependence. Both type and scale dependence may co-occur. In contrast, I refer to *type dependence only* as a setting where returns exogenously differ ($r_i \neq r_{i'}$ and $\frac{dr_i}{da_i} = 0$). Let there be a linear tax rate τ_K on capital gains $a_{R,i} \equiv a_i r_i(a_i)$. Suppose that utility is quasilinear in the consumption of final wealth. Utility maximization yields each household's Marshallian savings supply function $a_i(1-\tau_K, r_i(a_i); w_i)$ with $\frac{da_i}{dw_i} > 0$ and an indirect (present-value) utility $U(\tau_K; w_i)$. Define the elasticity of savings as $\zeta_i^{a,(1-\tau_K)} \equiv \frac{dlog(a_i)}{dlog(1-\tau_K)}$ and the capital gains elasticity as $\zeta_i^{a_R,(1-\tau_K)} \equiv \frac{dlog\left(a_{R,i}\right)}{dlog(1-\tau_K)}$. Without scale dependence, the return rates are fixed. Then, the two elasticities coincide $\tilde{\zeta}_i^{a,(1-\tau_K)} = \tilde{\zeta}_i^{a_R,(1-\tau_K)}$, where $\tilde{\zeta}_i$ indicates that the respective elasticity is evaluated at a fixed return rate. Under scale dependence, this is not the case. Let $\zeta_i^{a,r} \equiv \frac{dlog(a_i)}{dlog(r_i)}$ measure the responsiveness of savings to the rate of return. The novelty of this paper is to introduce scale dependence. The *own-return elasticity* $\varepsilon_i^{r,a} \equiv \frac{dlog[r_i(a_i)]}{dlog(a_i)}$ describes the extent of scale dependence. For simplicity, let $\tilde{\zeta}_i^{a,(1-\tau_K)}$, $\zeta_i^{a,r}$, and $\varepsilon_i^{r,a}$ be constant.[4]

Suppose that the utilitarian social planner wishes to maximize (steady-state) welfare at a given budget by optimally choosing the capital tax: $\underset{\tau_K}{max} \int_i \Gamma_i U(\tau_K; w_i)\, di$ subject to $\int_i \tau_K a_{R,i} di \geq \overline{E}$, where Γ_i is household i's Pareto weight, Γ_i is weakly decreasing in i, and $\int_i \Gamma_i di = 1$.[5] In the following, I use this basic framework to study capital taxation under scale dependence in contrast to the one under type dependence holding all the other primitives of the economy fixed (such as

[3] I microfound the notion of scale dependence, later, on a financial market with portfolio choice and financial knowledge acquisition. However, the findings carry over for any form of scale dependence (e.g., liquidity constraints).

[4] The assumption that $\varepsilon_i^{r,a}$ is constant over the population finds support in my empirical analysis of Section 3.

[5] Observe that both type and scale dependence feature return inequality. Whenever there is return inequality, the government wishes to levy a non-zero capital tax even if there is a nonlinear labor income tax available. That is, the zero-capital-taxation result (e.g., Atkinson and Stiglitz (1976), Judd (1985), and Chamley (1986)) breaks down. The intuition is that the presence of return inequality makes household heterogeneity two-dimensional. The government, then, uses the capital gains tax as an additional screening device (e.g., Saez (2002)).

the savings elasticities at a given rate of return). I establish five novel findings that I summarize in the following Proposition 1.

Proposition 1. *Compare the optimal capital gains tax in the presence of scale dependence to the capital gains tax in an economy with type dependence as the only source of return inequality.*

(*a*) *When expressed in terms of sufficient statistics, the conditions that describe the optimal capital gains tax with and without scale dependence are the same.*

(*b*) *Under scale dependence, an inequality multiplier effect increases the elasticity of capital income (relative to type dependence). This acts as a force for lower taxes.*

(*c*) *The optimal capital tax with scale dependence is either the same or lower than the capital optimal tax without scale dependence (type dependence only).*

(*d*) *Under scale dependence, a rise in capital taxes compresses the distribution of pre-tax returns. This compression effect comes along with the cost of lowering mean pre-tax returns.*

(*e*) *In the general equilibrium of the financial market of Section E, trickle-up externalities provide a force for a higher capital gains tax than in partial equilibrium.*

Part (*a*). The government's problem yields a Ramsey formula for the optimal capital tax (e.g., Diamond (1975)) in both settings, under scale dependence and without scale dependence (with type dependence only), $\frac{\tau_K}{1-\tau_K} = \frac{1}{\overline{\zeta}^{a_R,(1-\tau_K)}} I(\tau_K)$, where $\overline{\zeta}^{a_R,(1-\tau_K)} \equiv \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} \zeta_i^{a_R,(1-\tau_K)} di$ and $I(\tau_K) \equiv \mathbb{E}\left[\frac{(1-\Gamma_i)a_{R,i}}{\mathbb{E}(a_{R,i})}\right]$ measure the average elasticity of capital income and the observed capital income inequality, respectively. Irrespective of how returns form, the mean elasticity of capital income and the observed inequality serve as sufficient statistics. A correct knowledge of these measures is, therefore, enough to characterize the optimal capital tax which gives part (*a*) of Proposition 1.

Part (*b*). Without scale dependence, capital income under type dependence is linear in savings $a'_{R,i}(a_i)|_{r_i} = r_i$ and $a''_{R,i}(a_i)|_{r_i} = 0$. With scale dependence, the rate of return is endogenous. This makes capital gains convex in savings $a'_{R,i}(a_i) = r_i(a_i)$ and $a''_{R,i}(a_i) = r'_i(a_i) > 0$. Consider an individual i. In a setting with type dependence only, the individual is endowed with an investment skill that allows her to realize a return r_i. Her capital gains proportionally rise with her amount of investment. Off equilibrium, to obtain the same capital income as another individual i', she needs to increase her savings substantially. Under scale dependence, the individual has the same return rate r_i in equilibrium than without scale dependence. However, she can reach the capital income of individual i' more easily. Still, she needs to save more. At the same time, she can raise her rate of return to a higher level (in the financial market, by acquiring financial knowledge). This convexity boosts the savings and capital income elasticities, as I describe in the following.

Without scale dependence (with type dependence only), the average elasticity of capital income is equal to the savings elasticity $\overline{\zeta}^{a_R,(1-\tau_K)}|_{r_i} = \tilde{\zeta}_i^{a,(1-\tau_K)}$ for a given rate of return. With scale dependence, the savings elasticity needs to account for an endogenous return adjustment. Therefore, the savings elasticity and, accordingly, the average capital income elasticity are revised upwards $\overline{\zeta}^{a_R,(1-\tau_K)} = \Phi_i \overline{\zeta}^{a_R,(1-\tau_K)}|_{r_i}$, where $\Phi_i \equiv \frac{1+\varepsilon_i^{r,a}}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}} = (1+\varepsilon_i^{r,a})\sum_{n=0}^{\infty}(\zeta_i^{a,r}\varepsilon_i^{r,a})^n > 1$ measures the inequality multiplier effect. The size of the adjustment is proportional to the inequality multiplier effect Φ_i. The interpretation is straightforward: A tax cut increases a household's savings (when the substitution effect dominates the income effect). Under scale dependence, however, as savings increase, the rate of return rises as well. The higher rate of return increases the incentives to saves. In response, rates of return adjust, and so on. Φ_i captures this infinite loop of reactions that arises with scale dependence. As a result, savings and capital gains react more elastic to tax reforms. Since the optimal capital tax is inversely related to the mean capital gains elasticity, its upward adjustment provides a force for lower capital taxes. Proposition 1 (b) follows.[6]

Part (*c*). How does the presence of scale dependence affect optimal capital taxes? On the one hand, as described, scale dependence raises the observed capital gains elasticity, which reduces taxes in the optimum. On the other hand, the presence of scale dependence has the potential to amplify wealth inequality greatly. The optimal capital income tax is increasing in observed inequality. Through this channel, one would expect higher taxes that would reduce observed inequality. Therefore, I consider the following two comparative statics exercises.

Firstly, I compare the optimal capital income tax with scale dependence, τ_K, to the tax, denoted as $\tilde{\tau}_K$, one would obtain in a baseline economy with the same but exogenous distribution of returns (type dependence only). This exercise is, in principle, non-trivial, as the measure of inequality that determines the optimal tax may be endogenous to the underlying tax code $I(\tau_K)$. With constant elasticities, however, $I'(\tau_K) = 0$. Therefore, compared to an economy with type dependence that is observationally equivalent in terms of inequality ($I(\tau_K) = I(\tilde{\tau}_K)$), taxes are lower in the economy featuring scale dependence because the capital income elasticities are higher $\frac{\tau_K}{1-\tau_K} = \frac{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}{1+\varepsilon_i^{r,a}}\frac{\tilde{\tau}_K}{1-\tilde{\tau}_K}$.[7]

To demonstrate the quantitative importance of endogenous returns for optimal taxes, I calculate the optimal revenue-maximizing capital tax with and without scale dependence in Table 1. Set the elasticity of savings with respect to the rate of return equal to 0.5. Table 1 shows optimal tax rates for realistic combinations of $\tilde{\zeta}_i^{a,(1-\tau_K)}$ and $\varepsilon_i^{r,a}$. As usual, the larger the savings elasticity, the lower

[6]One can interpret this finding as a Le Chatelier principle for capital (see Samuelson (1948)). In the long run, since return rates adjust, capital responds more elastic than in the short run for fixed return rates.

[7]In the dynastic economy of Section C, I show that a similar logic applies to a linear wealth tax.

Own-Return Elasticity	Compensated Elasticity $\tilde{\zeta}_i^{a,(1-\tau_K)} = 0.25$	$\tilde{\zeta}_i^{a,(1-\tau_K)} = 0.5$	$\tilde{\zeta}_i^{a,(1-\tau_K)} = 1$
Baseline Model (No Scale Dependence): Exogenous Inequality in r_i			
$\varepsilon_i^{r,a} = 0$	80	67	50
Microfounded Model (Scale Dependence): Endogenous Inequality in r_i			
$\varepsilon_i^{r,a} = 0.1$	78	63	46
$\varepsilon_i^{r,a} = 0.25$	74	58	41
$\varepsilon_i^{r,a} = 0.5$	67	50	33
$\varepsilon_i^{r,a} = 1$	50	33	20

Table 1: Optimal Rawlsian Capital Tax Rate ($\zeta_i^{a,r} = 0.5$ and $\mathit{\Gamma}_i = 0$).

the optimal capital tax. The novel aspect of this paper is to have a non-zero own-return elasticity. As a benchmark, I consider $\varepsilon_i^{r,a} = 0$ in the first row (no scale dependence). The other rows differ by the magnitude of scale dependence. An own-return elasticity of 0.5, for instance, means that doubling the savings raises the rate of return accumulated over a lifetime by fifty percent. This amount of scale dependence leads to a reduction in the revenue-maximizing tax rate of more than 25% (17 percentage points). In the empirical section, I find a modest own-return elasticity of 0.1, which, nonetheless, reduces the optimal capital tax by six percent relative to the benchmark. The own-return elasticity that I retrieve from Fagereng et al. (2020) of 0.9 leads to a 50%-reduction of the optimal capital tax relative to a setting without scale dependence.

Alternatively, one can interpret these back-of-the-envelope calculations as the difference between the optimal capital tax and the tax set by a politician who wrongly assumes that the inequality he observes does not come from scale dependence (but from type dependence only). Altogether, even for a relatively small amount of scale dependence, the implications for the optimal tax rate are sizable.

Secondly, in Section G, I derive the optimal nonlinear capital gains tax in a life-cycle economy and show that the optimal tax remains unchanged when one introduces scale dependence, holding the primitives of the economy fixed. For given preferences and a fixed wage distribution, the introduction of scale dependence is completely neutral. The rise in inequality just cancels the increase in elasticities. Altogether, despite its potential to boost wealth inequality, scale dependence either reduces the optimal capital tax or is completely neutral (Proposition 1 (*c*)). One can read this result as a possible justification for why capital taxes (e.g., in the US) have not gone up, although capital income inequality has mounted. If this rise in inequality came from scale dependence, one should not tax more.

Part (*d*). Interestingly, the distribution of pre-tax returns is endogenous to the tax code. To see this, consider the variance of returns $\mathbb{V}(r_i)$ and a rise in the capital gains tax $d\tau_K > 0$. Then, under scale dependence ($\varepsilon_i^{r,a} > 0$), the variance of pre-tax returns declines $d\mathbb{V}(r_i) = -2\mathbb{V}(r_i)\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K} < 0$. In other words, the elasticity of the pre-tax return variance with respect to the retention rate is positive $\zeta^{\mathbb{V}(r),(1-\tau_K)} \equiv \frac{dlog[\mathbb{V}(r_i)]}{dlog(1-\tau_K)} > 0$. A rise in marginal taxes, therefore, reduces the pre-tax return inequality. However, this compression effect of returns comes along with the cost of diminishing mean pre-tax returns $d\mathbb{E}(r_i) = -\mathbb{E}(r_i)\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K} < 0$, which shows part (d) of Proposition 1. Thus, scale dependence gives rise to a new model-inherent trade-off for tax policy. On the one hand, a government that raises capital taxes can realize novel equity gains by reducing the pre-tax return inequality. On the other hand, there are novel efficiency costs from lowering the level of pre-tax returns.

For fully type-dependent rates of return, only the distribution of after-tax returns but not the pre-tax return distribution respond to the tax system. In the presence of scale dependence, capital taxes also affect the distribution of pre-tax returns. As a result, distributional responses of pre-tax returns provide a potential source for empirically identifying the magnitude of scale dependence. If all the return inequality came from type dependence, there should be no reaction of mean pre-tax returns and their variance to tax reforms. Whenever there is some scale dependence, one can observe such a response. As mentioned above, the strength of the reaction is, in this simple framework, proportional to the amount of scale dependence, measured by $\varepsilon_i^{r,a}$. In Section 4, I demonstrate the nature of different tax reforms affects these distributional responses.

Part (*e*). In Section E, I microfound the notion of scale dependence. On a financial market, households optimally choose their portfolio and the amount of information they wish to acquire. Wealthier households invest more and, therefore, have a higher incentive to acquire financial knowledge than poorer investors. As a result of their better knowledge, the former obtain higher rates of return than the latter households. Portfolio returns become scale-dependent. In general equilibrium, an investor's rate of return is not only positively associated with her portfolio size but also depends on others' investment decisions $r_i\left(a_i, \{a_i\}_{i'\in[0,1]}\right)$. The *cross-return elasticity* $\gamma_{i,i'}^{r,a} \equiv \frac{\partial log[r_i(\cdot)]}{\partial log(a_{i'})}$ measures the responsiveness of a household i's return to the amount of investment by another household i' (similar to the cross-wage elasticity in Sachs et al. (2020)). I show that, when costs of information acquisition are linear and everyone acquires knowledge, a change in the savings by a household i' leads to the same change the returns of any other household i in the percentage points $\gamma_{i,i'}^{r,a} = \frac{1}{r_i}\delta_{i'}^{r,a}$. Moreover, $\delta_{i'}^{r,a}$ is decreasing $a_{i'}$. It is positive for small values of $a_{i'}$ and negative for large ones. This situation features trickle-up forces, where a cut in the capital income tax of the rich shifts economic rents from the bottom to the top. The intuition is as follows. Absent of income effects, a tax cut on the rich's capital income income increases their portfolio size and financial knowledge. Accordingly, their returns rise ($\varepsilon_i^{r,a} > 0$). This channel is also present in partial equilibrium. However, in general equilibrium, the amount of aggregate information also rises, as the rich become more informed, and the value of private information declines. As a result, the reward for the relatively small amount of information the poor purchase declines, leading to lower return rates for them ($\delta_{i'}^{r,a} < 0$). The tax cut on the rich increases their wealth but reduces the resources of the poor.

The optimal capital tax in general equilibrium reads as $\frac{\tau_K}{1-\tau_K} = \frac{1}{\overline{\zeta}_i^{a_R,(1-\tau_K)}} \mathbb{E}\left[\frac{\left(1-\Gamma_i\left(1+\gamma_i^{r,(1-\tau_W)}\right)\right)a_{R,i}}{\mathbb{E}(a_{R,i})}\right]$, where $\gamma_i^{r,(1-\tau_W)} \equiv \int_{i'} \gamma_{i,i'}^{r,a} \zeta_{i'}^{a,(1-\tau_W)} di'$ summarizes general equilibrium welfare externalities. Suppose that cross-return elasticities average out such that $\int_{i'} \gamma_{i',i'}^{r,a} di' = 0$.[8] Then, one can show that the average capital gains elasticity, $\overline{\zeta}_i^{a_R,(1-\tau_K)}$, declines relative to the partial equilibrium. Moreover, $\gamma_i^{r,(1-\tau_W)} = \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' < 0$. Both the general equilibrium externalities and the adjustment of the wealth elasticity call for higher taxes.

For small general equilibrium forces ($\delta_{i'}^{r,a} \approx 0$ and $\tau_K^{GE} \approx \tau_K^{PE}$), one can use a first-order Taylor approximation to compare the optimal capital income tax in general equilibrium to the tax rate set by a politician who wrongly assumes that only partial equilibrium forces are present and sets a tax, τ_K^{PE}, that generates a capital income distribution for which the tax is optimal (as proposed in Rothschild and Scheuer (2013, 2016)). The general equilibrium tax rate is larger than the one in this self-confirming policy equilibrium $\tau_K^{GE} > \tau_K^{PE}$. Consequently, trickle-up forces call for more redistribution in general equilibrium. This result is intuitive because cutting taxes would shift resources from poor to affluent households and lower welfare.

One may think about this as a situation of rent-seeking, where the rich take away income from the poor. It is optimal for the government to tax these rents away (see Piketty, Saez, and Stantcheva (2014) and Rothschild and Scheuer (2016)). Of course, this does not mean that the capital tax should be higher or lower than in the setting with type dependence only (without scale dependence). This comparative static only compares capital taxes in partial and in general equilibrium. To evaluate whether the presence of endogenous return rates that result from rent-seeking should lead to more or less redistribution relative to a situation where return rates are exogenous, one needs a precise notion of the relative strength of partial and general equilibrium forces. In the empirical section, I attempt to disentangle these.

[8] In the empirical analysis (Section 3), I find some support for this assumption.

One can also interpret this result in connection with the integration of financial markets. As markets become internationally more connected, general equilibrium effects vanish ($\gamma_{i,i'}^{r,a} \to 0$). Foreign investors gain better access to a country's financial market. Vice versa, domestic investors can participate in foreign markets more easily when integration proceeds. As a result, domestic investors' impact on the return rates on the financial market is inversely related to the degree of integration. In this trickle-up economy, the optimal capital gains tax and, therefore, the level of redistribution declines with the international integration of financial markets.

2.2 Microfoundation, Extensions, and Discussion

In the following, I discuss the model's main assumptions, their generality, potential extensions, and policy implications of the framework.

Microfoundation. To begin, I describe the financial market of Section E as *one* potential microfoundation of scale dependence. I consider a repeated Grossman and Stiglitz (1980) financial market, where households optimally choose their portfolio consisting of a risk-free bond and a risky stock and acquire information about the stochastic fundamentals that drive the stock's payoff. In the rational expectations equilibrium, the stock price clears the market for individuals' portfolios, and the implied informativeness of the price is consistent with individuals' information acquisition.[9] I incorporate taxes into this market and demonstrate the functional form of own- and cross-return elasticities in a linear example. Moreover, I incorporate career effects and explicitly add type dependence.

Even though scale dependence arises, in this leading example, from information acquisition on a financial market, the exact source of scale dependence is in principle unimportant for tax policy. In partial equilibrium, only the magnitude of the own-return elasticity throughout the wealth distribution, $\varepsilon_i^{r,a}$, matters. To identify the direction of general equilibrium externalities, if at all present, the sources of scale dependence are relevant to the extent that they may enter differently into the cross-return elasticities, $\gamma_{i,i'}^{r,a}$. Also, notice that, in the conceptual framework, I do not assume away the joint presence of scale dependence and type dependence (e.g., entrepreneurial talent). Just as in reality, these phenomena can co-occur in my analysis.

Discussion and extensions. Besides, the message of the paper is not that taxes should be lower with return inequality than without. Instead, I analyze the policy implications of having a non-negligible responsiveness of return rates to savings. Lower taxes in the presence of scale

[9]For a more detailed exposition, I refer to Section E.

dependence do also not mean that the government should let wealth and return inequality grow indefinitely. There seems to be an upper bound on households' long-run return rates naturally limiting the amount of scale dependence and, thereby, the upward adjustment in capital income elasticities. In addition, the optimal capital tax rises with observed capital income inequality to combat rising inequality.

As already mentioned, most of the simplifying assumptions in this section are inessential for the main results. In the later sections, I show that those extend to income effects, the presence of labor taxes (Sections C and G), uncertain returns, and dynastic considerations (both Section G). For simplicity, I consider, in this section, a linear instead of a nonlinear capital gains tax (Section G). Alternatively, one can analyze a wealth tax as in Section C. As I show, all the insights hold for both capital gains and wealth taxation.

As noted by contributors to the literature (e.g., Guvenen et al. (2019)), a capital income and a wealth tax do not coincide when there is return heterogeneity. In the framework of Guvenen et al. (2019), type dependence generates return inequality between potentially liquidity-constrained entrepreneurs. A wealth tax can raise efficiency relative to a uniform capital gains tax as the former effectively levies a lower (higher) tax on capital incomes of individuals with a higher (lower) entrepreneurial talent and an exogenously higher (lower) rate of return. In their framework, return rates are independent from the amount of savings for unconstrained entrepreneurs and, given their calibration of the production function, even decreasing in the amount of savings for constrained ones. Therefore, the positive correlation between return rates and wealth in Guvenen et al. (2019) solely arises from type dependence. My model nests this type of return inequality. With scale dependence, there are additional efficiency gains because the lower tax on high-return individuals induces further efforts to increase their return rate. The focus of this paper is to study the effects of scale dependence on redistribution where efficiency is one (but not the only) important dimension.

Furthermore, I deal with the presence of other policies such as financial education consistent with the leading financial market example. Although such policies may be better suited to directly address return inequality, empirical evidence suggests, nonetheless, a residual amount of scale dependence governments cannot shut down. The reason is that these policies are also costly giving rise to a trade-off between equity (reduction in return inequality) and efficiency (rise in costs). Similarly, a government would face similar information acquisition costs if it provided a sovereign wealth fund open to everyone and large enough to absorb all private investment rents. Aside from information costs, this may give rise to other inefficiencies, such as agency frictions and diversification limits.

Even if these costs declined substantially, it seems unlikely that scale dependence would vanish. In the leading example of the financial market, the unpredictability of stock market returns may prevent the dissolution of scale dependence. Therefore, in this paper, I take as given existing inefficiencies that create a residual amount of scale dependence and analyze tax policy for this given amount of scale dependence.

Finally, the welfare weights may be endogenous to the amount of scale dependence. In the spirit of Saez and Stantcheva (2016), one may generalize the notion of social marginal welfare weights. For example, equity considerations may lead to even lower taxes when a given amount of return inequality comes from scale dependence as opposed to type dependence. In the latter case, rich individuals obtain higher rates of return than the poor, for instance, because of an inherent talent they received from their parents. Under scale dependence, individuals may inherit a sizable fortune which allows them, for example, to hire skilled financial advisers the poor cannot afford. More generally, they are gifted by their parents with the absence of frictions the poor have to face. However, at least partly the rich still need to pay a price to obtain higher rates of return than the poor, for example by taking effort. Thus, to some degree these higher returns reflect a fair compensation for costs the rich undertake. In that sense, scale dependence may reduce inequality concerns in a society, thus, lowering the optimal capital tax also from an equity perspective.

At the same time, political economy considerations may counteract this force. If the political power in a society is endogenous to an individual's wealth, the amplification of wealth inequality scale dependence causes may create a rich elite that either directly influences tax policy by running for a political office or indirectly by lobbying. For instance, as proposed by Saez and Zucman (2019), it may be desirable in the interest of sustaining democracy to set wealth taxes higher than the revenue-maximizing rate to prevent an "oligarchic drift". From this perspective, scale dependence may provide a rationale for higher capital taxes.

3 Empirical Analysis

In this section, I analyze the role of scale dependence for empirical analysis. First, I describe the conceptual issues that arise from scale dependence for the estimation of capital gains and savings elasticities and revisit estimates from the literature (Section 3.1). In this light, I, then, directly estimate own- and cross-return elasticities using panel data on US foundations (Section 3.2).

3.1 Empirical Implications of Scale Dependence

In this section, I describe the empirical implications of scale dependence for the estimation of capital gains and savings elasticities.

Conceptual description. To fix ideas, consider a tax reform $d\tau_K$ and abstract from income effects. Then, the percentage change in the capital gains of household i in period t is given by $\frac{da_{i,t}}{a_{i,t}} = -\tilde{\zeta}_{i,t}^{a,(1-\tau_K)} \frac{d\tau_K}{1-\tau_K} + \zeta_{i,t}^{a,r} \frac{dr_{i,t}}{r_{i,t}}$. Similarly, the change in the household's wealth, $a_{R,i,t} = a_{i,t} r_{i,t}$, reads as $\frac{da_{R,i,t}}{a_{R,i,t}} = -\tilde{\zeta}_{i,t}^{a_R,(1-\tau_K)} \frac{d\tau_K}{1-\tau_K} + \left(1 + \zeta_{i,t}^{a,r}\right) \frac{dr_{i,t}}{r_{i,t}}$.

This formulation immediately reveals the econometric implication of scale dependence for the estimation of *long-run* capital elasticities. In the presence of scale dependence, estimates from data that implicitly hold the return rate fixed ($dr_{i,t} = 0$) suffer from an omitted variable bias when trying to identify the long-run elasticities. Then, the estimation misses the adjustment of returns and the error term has a non-zero expectation, conditional on the covariates, violating a critical identifying assumption in empirical studies. The according point estimates are biased downward. In other words, wealth and capital income appear to be less responsive than they are in reality.

In the following, I describe three scenarios where this may be the case. Firstly, estimates may be biased when the empiricist does not correctly observe fluctuations in return rates. For data from a short time window, this is likely the case. In the short run, a household's return rate forms, for instance, conditional on her financial knowledge or advisers. In the longer term, she may react to tax reforms, e.g., by hiring other financial advisers, and the return rate adjusts.

Secondly, using data from tax records, the empiricist misses unrealized capital gains. These are not only but particularly relevant for households from upper parts in the wealth distribution, who, for instance, buy stocks or private equity and do not sell them. Therefore, the empiricist does not observe substantial parts of the adjustment in their capital income in response to a tax reform even if the data capture a long time. This problem also applies to housing, intangible properties, and other assets whose market value only reveals when being sold.

Another issue is the extrapolation of estimates from one to another group in the wealth distribution, even if they are unbiased. Portfolios and their flexibility differ significantly across the population. Households from low parts of the wealth distribution mostly hold cash and cannot participate in the stock market. Median families hold mostly housing. For wealthy households, financial and business assets are pervasive. Therefore, one cannot infer estimates of capital income elasticities from the poor to the rich and vice versa. To overcome these issues, one may directly estimate own- and cross-return elasticities along the wealth distribution. I approach this empirical challenge in Section 3.2.

Relation to the empirical literature. Now, I summarize two strands of the empirical literature bearing these issues in mind. The first one regards the estimation of the capital income elasticity with respect to the capital gains tax. In the second strand of the literature, contributors estimate the elasticity of capital to wealth taxes. The number of studies that try to address scale dependence is limited. This does, of course, not mean that the other estimates are wrong, but their scope of application depends on the nature of policies under consideration.

Contributors to the literature on the capital income elasticity, starting from Feldstein, Slemrod, and Yitzhaki (1980), employ microdata and time-series mostly from the US. Their focus lies on the estimation of *realization elasticities* (for recent contributions, see Bakija and Gentry (2014), Dowd, McClelland, and Muthitacharoen (2015), and Agersnap and Zidar (2020)). The authors distinguish between transitory and permanent responses. Permanent responses seem to be more relevant for long-run tax policy. However, to apply to long-run capital taxation, estimates also need to control for scale dependence in return rates. Existing studies may not capture them because they are from a short-time window and only include capital gains realizations.

Unlike the sizable research on the elasticity of taxable income, only a few studies have, so far, attempted to estimate the elasticity of capital with respect to wealth taxes. Zoutman (2015) studies the impact of a capital tax reform on wealth accumulation in the Netherlands, noting that the portfolio composition changes over time and responds to the tax reform. However, the data only include cash returns (e.g., dividends and interest), thereby lacking a measure of actual returns. In their analysis of Swiss time-series and microdata, Brülhart, Gruber, Krapf, and Schmidheiny (2016) omit adjustments in individual return rates. Seim (2017) provides evidence of bunching at exemption thresholds in Sweden. Whereas being suited for identifying avoidance and evasion responses, such estimates need to be interpreted locally for the respective wealth group and may not represent real responses in the long run (see Kleven (2016)). In Denmark, Jakobsen, Jakobsen, Kleven, and Zucman (2020) estimate the wealth elasticity in a difference-in-difference setup. Since the Danish wealth tax only applies to wealthy households in the observation period, the estimates are not representative of the entire population. To sum up, this literature pays closer attention to unrealized capital gains, which is natural, given its objective to estimate the wealth elasticity. However, without knowing the amount of scale dependence, the estimates are not readily generalizable to long-run wealth elasticities across the wealth distribution.

3.2 Estimation of Own- and Cross-Return Elasticities

In the following, I propose a direct estimation of own- and cross-return elasticities that can be used to adjust the elasticity of capital to scale dependence. As I demonstrate in Sections 2, C, and G, the adjustment of this sufficient statistic is important for tax policy.

Foundations data. I use the publicly available panel data on US foundations that annually report their wealth and income to the IRS in the 990-PF form. The stratified random sample covers approximately 10% of the foundation population. This procedure is similar to Piketty (2014), who uses pooled returns data of US universities. The micro-files on foundations cover the years 1986 to 2016. They include market-valued wealth levels, portfolio compositions, and capital income. All observations are on an individual level.

The foundation data set has three main advantages. Firstly, it allows me to follow the relation between return rates and wealth on an individual (foundation) level over a long period. Secondly, although foundations are institutional investors who potentially behave differently on the financial and non-financial markets, they may serve as a reasonable proxy for wealthy investors. Their portfolios' size is similar, and their assets are also partly shifted to legal entities instead of private bank accounts. Thirdly, the data set contains both realized and unrealized capital gains, and foundations explicitly report donations and withdrawals.

The main disadvantage of the data set is its limited generalizability to household behavior. The average foundation has a substantially larger endowment than the average household, and, even conditional on the same wealth level, investment behavior may differ. However, one may argue that foundations provide a reasonable proxy for the rich with a similar portfolio size who partly shifts their assets to these entities. Nonetheless, one should be cautious when interpreting the findings in the context of households.

As in Fagereng et al. (2020) and Bach et al. (2020), one can directly calculate the real investment return on wealth of foundation i during period h, $r_{i,h}$, as the market-value capital income (both

Wealth Group I_g	g	Relative Group Size	Wealth Level Mean	Return Rate Mean	Return Rate Std Dev
Below $100k	1	7.9%	40 574	2.7%	(0.072)
$100k to $1m	2	20.6%	416 464	4.9%	(0.076)
$1m to $10m	3	24.6%	3 708 144	5.1%	(0.084)
$10m to $100m	4	39.9%	31 065 438	5.0%	(0.088)
$100m to $500m	5	5.7%	197 396 730	5.4%	(0.091)
$500m to $5bn	6	1.1%	1 207 748 745	5.8%	(0.090)
Above $5bn	7	0.1%	10 238 369 724	5.7%	(0.097)

Table 2: Summary Statistics (Observations: $N = 254\,570$)

realized and unrealized) divided by the average invested capital in that period. Denote foundation i's assets at market value at the beginning of year h as $a_{i,h}$. All the observations are in 2016 dollars. By construction of the empirical specifications below, I only use foundation-year observations with positive beginning-of-year assets. Moreover, to avoid outliers, I exclude foundation-year observations with return rates above 25% and below −25%.[10] Moreover, I drop foundation observations with zero wealth at the beginning of a year. As in Saez and Zucman (2016), I classify foundations by their market-value wealth at the beginning of each year into wealth groups $g = 1, ..., 7$ (index set I_g). In Table 2, I display descriptive statistics for these different wealth groups.

The first three (four) wealth groups approximately capture the bottom 50% (90%) of foundations. The last two groups cover the top 1% and the top 0.1%, respectively. Foundations achieve a median return rate of 4.9% with an median portfolio size of $6 978 721. There is a substantial degree of heterogeneity. Foundations differ in their endowment size (wealth inequality) and in their investment returns (return inequality). Whereas small foundations (below $100k) attain an annual return of 2.7%, the top 1% foundations gain 5.8% on their investments. A 1% increase in the endowment size is associated with a reduced-form rise in the annual return rate of 0.2%. Notice that the average foundation is substantially wealthier than the typical household. At the same time, their return rates are comparable. Therefore, the amount of scale dependence is likely to be underestimated in the data. The estimates of the own-return elasticity can be considered conservative.

Estimation of own-return elasticities. To disentangle the role of type and scale dependence for this return inequality, I utilize the data's panel structure in the following. Like Fagereng et al.

[10] If one leaves out foundation-year observations with return rates below the 2.5 and above the 97.5 percentile, the results will be similar. The uncut sample features a kurtosis above 100 000 which is far beyond any threshold proposed in the literature for evaluating outliers and fat tails (e.g., see Kline (2015)). After cutting the sample in the proposed manner, the kurtosis drops to 3.7, thus resembling a normal distribution's tail behavior.

	Constant Returns to Scale			Incr./Decr. Returns to Scale
	(1)	(3)	(4)	(2)
ε	0.0023***	0.0022***	0.0001***	0.0027***
	(0.0004)	(0.0004)	(0.0000)	(0.0005)
ε_2				0.0002
				(0.0001)
ε_3				0.0002
				(0.0002)
ε_4				0.0001
				(0.0002)
ε_5				−0.0002
				(0.0002)
ε_6				−0.0003
				(0.0002)
ε_7				−0.0008
				(0.0006)
Individual FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Observations	254 570	254 570	254 570	254 570

Table 3: Own-Effects Regressions; Standard Errors (in Parentheses) Clustered by Foundation; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

(2020), I regress real return rates on *beginning-of-year* net wealth

$$log\left(1 + r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + f_i + f_h + u_{i,h}, \tag{1}$$

where f_i and f_h are individual and time fixed effects. ε measures scale dependence, whereas f_i captures the amount of type dependence. Therefore, individual-specific time variation in wealth identifies scale dependence that arises from any direct or indirect source (e.g., portfolio choice, financial information, stock market participation costs, or liquidity). For instance, donations or withdrawals trigger such time variation in portfolio size.

There may be nonlinearities in scale dependence. In the example of the financial market in Section E.2.2, there are decreasing returns to scale. The own-return elasticity decreases with wealth. To capture these nonlinearities, I estimate an alternative specification

$$log\left(1 + r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + \sum_{g'=2}^{7} \varepsilon_{g'} \cdot log\left(a_{i,h}\right) \cdot D_{g_{i,h},g'} + f_i + f_h + u_{i,h}, \tag{2}$$

where $D_{g_{i,h},g'}$ is a dummy variable indicating a foundation i's affiliation to group g' in period h.

In Table 3, I report the estimated coefficients of specifications (1) and (2). Specifications (1) and

(2) reveal a highly significant amount of scale dependence. Doubling a foundation's endowment, raises its annual return rate by 0.23 percentage points (period-h own-return semi-elasticity). There is no evidence for increasing or decreasing returns to scale. Interestingly, for high foundations sizes, the point estimates of (2) show (slightly non-significant) decreasing returns to scale that would be in line with the financial market example in Section E.2.2. Whereas the specification cannot confirm the parametrization in the example, it does neither reject it.

In specifications (3) and (4), I replace log net wealth in (1) by foundation's group affiliation and percentile in the wealth distribution both based on foundations' beginning-of-year net wealth:

$$log\,(1 + r_{i,h}) = \varepsilon \cdot g_{i,h} + f_i + f_h + u_{i,h}, \tag{3}$$

and

$$log\,(1 + r_{i,h}) = \varepsilon \cdot p_{i,h} + f_i + f_h + u_{i,h}, \tag{4}$$

where $g_{i,h}$ and $p_{i,h}$ measure foundation i's wealth group affiliation and percentile in period h. Again, there is significant scale dependence. Including lagged foundation wealth into (1) does not change the results qualitatively. Instrumenting foundation wealth in (1) with three-year lagged donations yields the same results. By estimating (1) separately for boom and bust years (1990, 2001, 2008, and 2009), one can show that scale dependence is driven by boom years. Large foundations realize large capital gains (losses) during boom (bust) years because they take more risks than small ones (higher return variance).

Now, I translate the scale dependence estimated with (1) into a value for the life-time own-return elasticity $\varepsilon_i^{r,a}$. Multiply the estimate of (1) by $\frac{1+r_{m,h}}{r_{m,h}}$, where $r_{m,h} = 4.9\%$ is the median return rate, to get an estimate of the period-h own-return elasticity of a representative foundation ($\widehat{\varepsilon_{i,h}^{r,a}} \approx 0.05$). To compute the life-time own-return elasticity, consider the compound return rate earned by that household $R_m = (1 + r_{m,h})^H - 1$. Accordingly, one obtains an expression for the life-time own-return elasticity $\widehat{\varepsilon_i^{r,a}} = \frac{H \cdot (1+r_{m,h})^{H-1}}{R_{m,h}} \frac{dlog(1+r_{i,h})}{dlog(a_{i,h})}$. For $r_{m,h} = 4.9\%$ and $H = 30$, this yields an estimate of $\widehat{\varepsilon_i^{r,a}} \approx 0.1$. As we have seen, incorporating this conservatively estimated effect into tax policy leads to a notable adjustment of the optimal capital tax.

By comparing the estimate of (4) to the one in Fagereng et al. (2020), one can immediately see that the predictions from the foundations' data set may severely understate the amount of scale dependence among households. Based on the wealth distribution in Norway and the estimated scale dependence in Fagereng et al. (2020), I calculate an estimate of the life-time own-return elasticity of $\widehat{\varepsilon_i^{r,a}} \approx 0.9$ in their data set, which is substantially higher than the estimate obtained here.[11]

One can explain the different estimates with the difference between the foundation data the Scandinavian population data in Fagereng et al. (2020) and Bach et al. (2020). When extrapolating the US foundation data to a country's population, say US households, two important notes of caution are, thus, in order. Firstly, the foundations' portfolios are substantially larger. For example, in the SCF data in Section 4, having the median foundation wealth would put a US household into the 96th percentile in the wealth distribution. Therefore, the bottom half of the foundations may better proxy the population than the whole foundation distribution.

Secondly, a household and a foundation with the same net wealth potentially behave very differently. Consider a household and a foundation with $ 10 000. The household has comparably small savings such that liquidity constraint may play an important role. Simultaneously, the foundation may not face such constraints if it is funded by an affluent donor who endows several charities and bails each of these foundations out whenever necessary. Altogether, the different datasets lead to distinct estimates of scale dependence. In any case, the implications of scale dependence for tax policy are sizable.

Estimation of cross-return elasticities. Recall that, in general equilibrium, a household's return rate $r_i\left(a_i, \{a_{i'}\}_{i'\in[0,1]}\right)$ depends not only on its amount of savings but also on those of others. A change in the savings by household i' also affects household i's return

$$\frac{d\left(1+r_{i,h}\right)}{1+r_{i,h}} = \varepsilon_{i,h}^{1+r,a} \cdot \frac{da_{i,h}}{a_{i,h}} + \int_{i'} \gamma_{i,i',h}^{1+r,a} \cdot \frac{da_{i',h}}{a_{i',h}} di'.$$

To bring this formulation closer to the data, consider the discrete counterpart

$$\frac{d\left(1+r_{i,h}\right)}{1+r_{i,h}} = \varepsilon_{i,h}^{1+r,a} \cdot \frac{da_{i,h}}{a_{i,h}} + \sum_{i'} \gamma_{i,i',h}^{1+r,a} \cdot \frac{da_{i',h}}{a_{i',h}}.$$

In the following, I estimate the magnitude of general equilibrium externalities (for each wealth group). To be able to identify cross effects, I impose more structure on these externalities. I assume that they are constant over time ($\gamma_{i,i',h}^{1+r,a} = \gamma_{i,i',h}^{1+r,a}$) and multiplicatively separable $\gamma_{i,i',h}^{1+r,a} = \frac{1}{1+r_{i,h}}\delta_{i',h}^{r,a}$, as in the financial market example (Section E.2.2). Moreover, let general equilibrium externalities

[11] Using the wealth distribution reported in Table 1A of Fagereng et al. (2020), I regress the household percentile on log wealth ($\widehat{\frac{dp_{i,h}}{dlog(a_{i,h})}} = 0.1443$). Then, note that $\widehat{\frac{dr_{i,h}}{dlog(a_{i,h})}} = \widehat{\frac{dr_{i,h}}{dp_{i,h}}}\widehat{\frac{dp_{i,h}}{dlog(a_{i,h})}}$, where $\widehat{\frac{dr_{i,h}}{dp_{i,h}}} = 0.1383$ (see Table 9 in Fagereng et al. (2020)), to obtain an estimate for the period-h own-return semi-elasticity. Finally, for $r_{m,h} = 3.2\%$ (reported in Table 3 of Fagereng et al. (2020)) and $H = 30$, I obtain a period-h own-return elasticity of $\widehat{\varepsilon_{i,h}^{r,a}} \approx 0.6$ and a life-time own-return elasticity of $\widehat{\varepsilon_{i}^{r,a}} \approx 0.9$. For $r_{m,h} = 5.6\%$, as in the SCF data in Section 4, $\widehat{\varepsilon_{i,h}^{r,a}} \approx 0.4$ and $\widehat{\varepsilon_{i}^{r,a}} \approx 0.7$.

be similar within a wealth group $\delta^{r,a}_{i',h} \approx \delta^{r,a}_{g',h}$ for all $i' \in I_{g'}$ and let $\delta^{r,a}_{g',h}$ be small ($\delta^{r,a}_{g',h} \approx 0$). In the estimation, I verify the latter assumption. Define the mean return in wealth group g as $\mathbb{E}_g\left(r_{i,h}\right)$. Then one can write the effect on returns as

$$\frac{d\left(1+r_{i,h}\right)}{1+r_{i,h}} = \varepsilon^{1+r,a}_{i,h} \cdot \frac{da_{i,h}}{a_{i,h}} + \sum_{g'=1}^{7} \delta^{r,a}_{g',h} \cdot \sum_{i' \in I_{g'}} \frac{da_{i',h}}{a_{i',h}} \cdot \frac{1}{1+\mathbb{E}_g\left(r_{i,h}\right)} + u_{i,h}$$

with a bias term

$$u_{i,h} \equiv \sum_{g'=1}^{7} \frac{1}{\left(1+r_{i,h}\right)\left(1+\mathbb{E}_g\left(r_{i,h}\right)\right)} \sum_{i' \in I_{g'}} \left[\left(1+\mathbb{E}_g\left(r_{i,h}\right)\right)\left(\delta^{r,a}_{i',h} - \delta^{r,a}_{g',h}\right) + \left(\mathbb{E}_g\left(r_{i,h}\right) - r_{i,h}\right)\delta^{r,a}_{g',h}\right] \frac{da_{i',h}}{a_{i',h}}.$$

For small cross effects ($\delta^{r,a}_{g',h}$) and return rates ($r_{i,h} - \mathbb{E}_g\left(r_{i,h}\right)$) and similar cross-effects in each wealth group ($\delta^{r,a}_{i',h} \approx \delta^{r,a}_{g',h}$), the bias term becomes negligible $u_{i,h} \approx 0$. Therefore, I specify the econometric model by augmenting (1) with cross effects

$$log\left(1+r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + \sum_{g'=1}^{7} \delta_{g'} \cdot \overline{log\left(a_{g',h}\right)} \cdot g_{i,h} + f_i + f_h + u_{i,h} \tag{5}$$

and, controlling for group-specific effects,

$$log\left(1+r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + \beta \cdot g_{i,h} + \sum_{g'=1}^{7} \delta_{g'} \cdot \overline{log\left(a_{g',h}\right)} \cdot g_{i,h} + f_i + f_h + u_{i,h} \tag{6}$$

where, again, $g_{i,h}$ indicates foundation i's group affiliation and $\overline{log\left(a_{g',h}\right)} \equiv \sum_{i' \in I_{g'}} log\left(a_{i',h}\right)$ measures the wealth level of group g in period h.

There are two sources for identifying $\delta_{g'}$: movements in the groups' wealth levels and foundations' mobility between wealth groups. Changes in the foundations' group affiliation arise from donations, withdrawals, and investment returns in the past. In Table 4, I display the amount of inter-group mobility. As the diagonal of this mobility matrix reveals, there is a substantial group persistence (96% of observations). The majority of foundation mobility is between adjacent wealth groups. There is slightly more upward than downward mobility. Overall, 9 048 foundation-year group movements identify the inter-group cross effects.

In Table 5, I show the coefficients estimated from (5) and (6). The estimated amount of scale dependence remains relatively stable. Moreover, there are statistically significant cross-effects. The estimates reveal no clear relationship between $\delta_{g'}$ and g'. In the financial market example of Section E.2.2, this relation would be negative. However, the sizes of the significant coefficients are, from an economic point of view, negligible. The estimates justify using small general equilibrium forces

	$g_{i,h}=1$	$g_{i,h}=2$	$g_{i,h}=3$	$g_{i,h}=4$	$g_{i,h}=5$	$g_{i,h}=6$	$g_{i,h}=7$
$g_{i,h-1}=1$	15 370	599	70	19	1	1	0
$g_{i,h-1}=2$	1 073	43 217	1 239	30	1	1	0
$g_{i,h-1}=3$	29	1 106	52 820	1 553	13	2	0
$g_{i,h-1}=4$	6	10	1 035	89 911	1 151	16	0
$g_{i,h-1}=5$	0	0	1	718	12 519	218	1
$g_{i,h-1}=6$	0	0	0	1	134	2 536	14
$g_{i,h-1}=7$	0	0	0	0	0	6	216

Table 4: Inter-Group Mobility (Observations: $N = 225\,637$)

($\delta_{i,t}^{r,a} \approx 0$) in the comparative statics (Sections 2 and C) and validate the identifying assumption that $u_{i,h} \approx 0$ (for $\delta_{i',h}^{r,a} \approx \delta_{g',h}^{r,a}$).

If at all, the estimates of δ_7 are economically relevant. As group 7 represents the top 0.1% of foundations, this indicates the presence of negative externalities from the top hinting at trickle-up forces in the general equilibrium financial market. Moreover, using the estimated coefficients from specification (6), a simple Wald test does not reject the hypothesis that $\int_{i'} \gamma_{i',i'}^{r,a} di' = 0$ at the 5% level, which is in line with the assumption in the theoretical part.

To account for potential group-specific nonlinearities in cross effects, I also estimate

$$log\left(1+r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + \sum_{g'=1}^{7} \sum_{g''=1}^{6} \delta_{g',g''} \cdot \overline{log\left(a_{g',h}\right)} \cdot D_{g_{i,h},g''} + f_i + f_h + u_{i,h} \tag{7}$$

and

$$log\left(1+r_{i,h}\right) = \varepsilon \cdot log\left(a_{i,h}\right) + \sum_{g''=1}^{6} D_{g_{i,h},g''} + \sum_{g'=1}^{7} \sum_{g''=1}^{6} \delta_{g',g''} \cdot \overline{log\left(a_{g',h}\right)} \cdot D_{g_{i,h},g''} + f_i + f_h + u_{i,h}, \tag{8}$$

where $D_{g_{i,h},g''}$ is a dummy variable equal to one if foundation i's period-h group affiliation $g_{i,h}$ is g''. As in (5) and (6), the estimated cross effects ($\delta_{g',g''}$) are economically small. Moreover, the estimates do not reveal noteworthy nonlinearities in cross effects. Therefore, I abstain from reporting them separately.

Altogether, I find a statistically significant and economically meaningful amount of scale dependence. The preferred estimate leads to an own-return elasticity of 0.1. Using the statistics reported in Fagereng et al. (2020), I retrieve an estimate of 0.9 in their data set. In both cases, the resulting adjustment of capital elasticities and the implications for tax policy are quantitatively important. The cross-effects estimates are statistically significant but economically unimportant, suggesting either no or only small general equilibrium externalities. Some of the cross-effects es-

	Constant Returns to Scale		Constant Returns to Scale	
	(5)		(6)	
ε	0.0008*		0.0025***	
	(0.0004)		(0.0004)	
δ_1	−0.0005*	$\times 10^{-3}$	0.0007***	$\times 10^{-3}$
	(0.0003)	$\times 10^{-3}$	(0.0002)	$\times 10^{-3}$
δ_2	0.0006	$\times 10^{-4}$	−0.0031***	$\times 10^{-4}$
	(0.0010)	$\times 10^{-4}$	(0.0009)	$\times 10^{-4}$
δ_3	0.0030***	$\times 10^{-4}$	0.0049***	$\times 10^{-4}$
	(0.0007)	$\times 10^{-4}$	(0.0007)	$\times 10^{-4}$
δ_4	−0.0023***	$\times 10^{-4}$	−0.0038***	$\times 10^{-4}$
	(0.0007)	$\times 10^{-4}$	(0.0008)	$\times 10^{-4}$
δ_5	0.0076	$\times 10^{-4}$	0.0306***	$\times 10^{-4}$
	(0.0049)	$\times 10^{-4}$	(0.0055)	$\times 10^{-4}$
δ_6	0.0036***	$\times 10^{-3}$	0.0024**	$\times 10^{-3}$
	(0.0011)	$\times 10^{-3}$	(0.0011)	$\times 10^{-3}$
δ_7	−0.0022***	$\times 10^{-2}$	−0.0037***	$\times 10^{-2}$
	(0.0007)	$\times 10^{-2}$	(0.0007)	$\times 10^{-2}$
Individual FE	Y		Y	
Time FE	Y		Y	
Observations	254 570		254 570	

Table 5: Cross-Effects Regressions; Standard Errors (in Parentheses) Clustered by Foundation; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

timates seem to be in line with the specified general equilibrium financial market model. More research is needed to assess in how far the estimates from foundations apply to household data and whether general equilibrium externalities are present.

4 Quantitative Analysis

Now, I document the substantial amount of return inequality in the US, using household-data from the Survey of Consumer Finances (2016). Then, I use the type and scale dependence estimates in the previous section and match the statistical model (1) to the cross-section of return rates in the SCF. I show that the statistical model successfully explains the observed inequality in return rates. Using the SCF data, I consider a set of tax reforms that alter the progressivity of the current US tax code and study their distributional and revenue incidence. For a calculation of optimal tax rates, I refer to Sections 2 and G.

4.1 Return Inequality in the US

Data. I extract the household-level asset data from the SCF for 2016 provided by Kuhn, Schularick, and Steins (2020). The representative sample contains detailed information on household wealth, portfolio composition, and demographic characteristics. For simplicity, I exclude around 16% of households, who report non-positive net wealth. I define net wealth as the market value of all financial and non-financial assets net of the value of total debt. Since income from pension funds and life insurance is exempt from capital taxation, I exclude these assets from the wealth concept.

Construction of return rates. In the following, I construct household-level return rates. First, I calculate portfolio shares for each household. Figure 1 displays the relationship between households' position in the wealth distribution and their portfolios' composition. There is a substantial amount of heterogeneity in the portfolio composition. Housing and other real estate is the most prevalent asset throughout the wealth distribution. Its importance is, however, declining in household wealth. Similarly, the share of liquid assets is considerably higher for households from lower parts of the wealth distribution. The opposite holds for public and private equity. Business assets and stocks are increasingly important throughout the wealth distribution. Debt is more prevalent for the poor and the middle class than for the rich. Bond holdings (interest-bearing) only play a minor role.

Now, using these asset shares, I construct household-level return rates. Formally, I calculate household i's period-h gross return rate as $r_{i,h} = \sum_a \varsigma^a_{i,h} r^a_h$, where $\varsigma^a_{i,h}$ is the share of asset a in the household's portfolio and r^a_h is the asset's annual return rate. For simplicity, I assume uniform

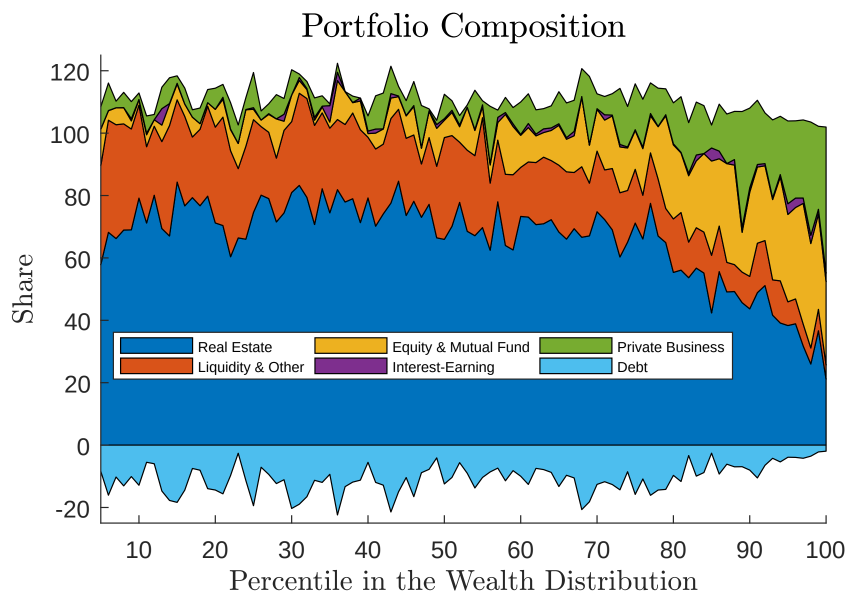

Figure 1: Asset Portfolio Composition

return rates, conditional on a given asset a.[12] This procedure will likely underestimate the actual degree of return inequality if return rates within an asset class rise with household wealth. I use each asset's 30-year average return rate in the US (1986-2015) from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019). To obtain real portfolio returns, I subtract the 30-year average inflation in the US, based on the CPI.

In Figure 2, I show the return rates for each percentile in the wealth distribution. The graph reveals a substantial return inequality throughout the wealth distribution. When moving from the lowest to the highest wealth percentile, households' annual return rates differ by more than 4.5 percentage points. This heterogeneity is remarkable given that I imposed uniform return rates within asset classes and omitted the poorest households (16%) with zero or negative net wealth.

4.2 Statistical Model

Model description and parametrization. I use the statistical model (1) estimated in the previous section from the foundation data to explain the observed cross-section of household returns. This exercise can be interpreted as an indirect test of the underlying data-generating process.

[12] The assumption of uniform return rates for a given asset class is in line with the microfoundation in Section E. There, return inequality arises from heterogeneous portfolio choices by households that differ in their financial knowledge.

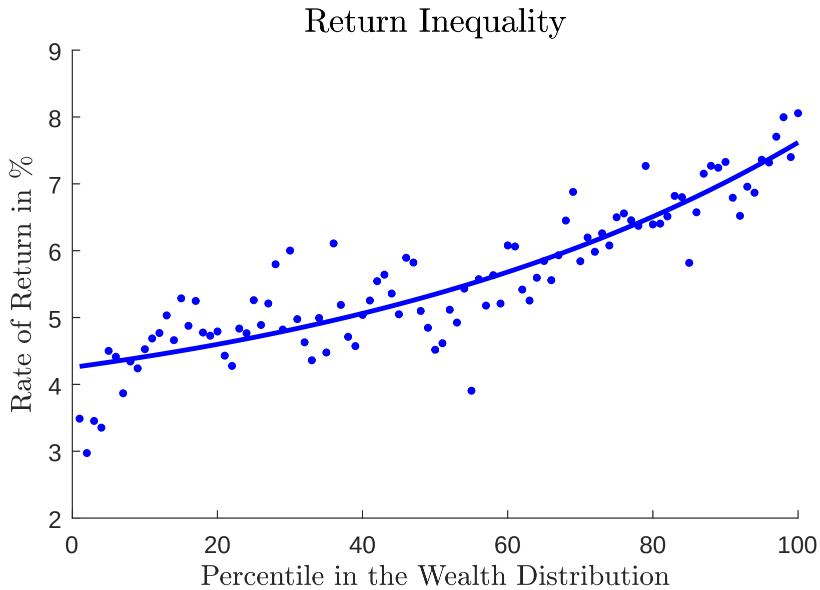

Figure 2: Pre-Tax Return Rates and Percentile in the Wealth Distribution

Moments	Model	Data
Mean	0.0556	0.0556
Median	0.0555	0.0556
Standard Deviation	0.0217	0.0216

Table 6: Moments of the Return Distribution (Model vs. Data)

Recall that the statistical model describing the relationship between return rates and wealth is given by $r_{i,h} = a_{i,h}^{\hat{\varepsilon}} exp\left(\hat{f}_i + \hat{f}_h\right) exp\left(\hat{u}_{i,h}\right) - 1$, where $\hat{\varepsilon} = 0.0023$ measures the amount of scale dependence and $\hat{u}_{i,h} \sim \mathcal{N}\left(0, \hat{\sigma}_u\right)$. I denote the amount of type dependence as $\hat{\upsilon}$. In the fixed-effects regression (1), this type dependence is hidden in the correlation between the individual fixed effects and the portfolio size. To recover $\hat{\upsilon}$, I regress the foundations' fixed effects on the foundations' mean log asset level, giving an estimate of $\hat{\upsilon} = 0.0011$. Using this statistical model, I construct the model-implied return rate for each household. Finally, I choose $\hat{\sigma}_u = 0.02$ and $\hat{f}_h = 0.0104$ to match the cross-sectional mean and standard deviation of return rates.

Model fit. As demonstrated in Table 6, the statistical model successfully explains key moments of the distribution of return rates. In Figure 3, I compare the distribution of return rates in the SCF to those constructed from the model. The Left Panel of the figure displays the cross-section of return rates. Although the foundations data from which I estimate the amount of scale and type dependence in the statistical model are quite different from the household data, the model-implied

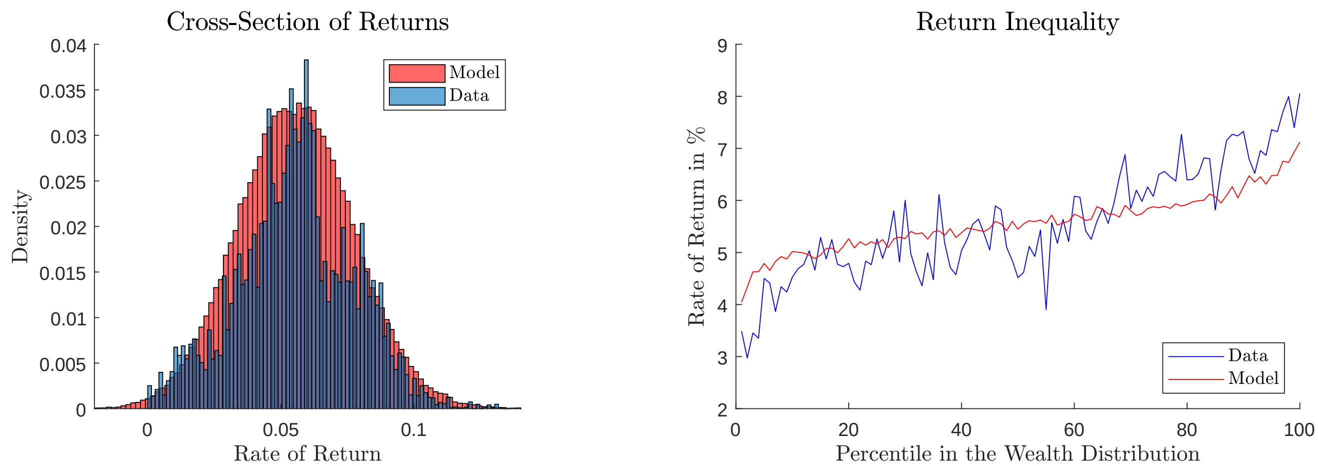

Figure 3: Left Panel: Cross-Section of Return Rates (Model vs Data); Right Panel: Return Inequality (Model vs Data)

cross-section of returns looks very similar to the actual one.

In addition to fitting the unconditional distribution of return rates, the model matches the observed degree of return inequality. In the Right Panel of Figure 3, I compare the model-implied return inequality to the observed inequality (Figure 2). As for the unconditional returns, the model performs well in explaining the dispersion in return rates throughout the wealth distribution. If at all, the model overestimates return rates at the bottom and underestimates the rich's returns. Combined with the fact that the construction of return rates in the data should understate the amount of return inequality, the statistical model's amount of scale dependence appears conservative.

4.3 Incidence Analysis

To gain some notion of the quantitative importance of scale dependence, I study the incidence of several tax reforms that change the current US tax code's progressivity on the level and dispersion of return rates, the capital gains tax base, and revenues. In the following simulations, I employ a sufficient statistics approach. I take households' savings decisions at the current US tax code observed in the SCF, perturb the capital gains tax, and simulate the counterfactual choices using a set of reduced-form elasticities.

Ignoring nonlinearities, the long-term capital gains tax in the US is currently $\tau_K = 0.15$. Observe that the government taxes realized capital gains only. The preceding computation of return rates includes both realized and unrealized capital gains. To overcome this issue, I construct a realization share of returns (i.e., the ratio between realized and overall capital gains) to match the annual amount of realized capital gains in the US (more than $700 billion). I assume, as a benchmark, that this realization share is uniform across the population.

Moreover, suppose that the elasticities are constant in the population. As in the back-of-the-envelope calculations, let $\zeta_{i,h}^{a,r} = 0.5$ and $\tilde{\zeta}_{i,h}^{a,(1-\tau_K)} = 0.5$. I scale up the period-own-return elasticity to $\varepsilon_{i,h}^{r,a} = 0.2$, a value between the conservative estimate obtained from the foundation data and the large one I retrieve from Fagereng et al. (2020). I compare the resulting incidence to an exogenous return benchmark (no scale dependence $\varepsilon_{i,h}^{r,a} = 0$). In the Figures in Appendix B.2, I also display simulations using the estimate of Fagereng et al. (2020) ($\varepsilon_{i,h}^{r,a} = 0.4$). Since general equilibrium externalities appear to be small in the empirical analysis of Section 3, I neglect them. Moreover, I hold the capital gains realization share fixed.

Tax reforms. I consider three types of tax reforms. The first type of tax reform (Reform 1) raises the capital gains tax for all individuals above a certain wealth level holding the rest of the tax code fixed. Reform 2 decreases capital taxes below some wealth threshold. The third reform (Reform 3) increases taxes above a threshold and reduces those below the threshold. Thus, the third reform is a combination of the first and second reform. Notice that all three reforms raise the progressivity of the current US tax code.

Distributional effects. Now, I study the distributional responses (i.e., the response of mean pre-tax returns and their dispersion) triggered by the three types of tax reforms more carefully. Figure 4 demonstrates the response of mean pre-tax returns (Left Panel) and their standard deviation (Right Panel) to the three sets of tax reforms depending on the respective wealth threshold (horizontal axes). The blue lines depict the initial mean and standard deviation observed under the current US tax code. They also represent a situation without scale dependence ($\varepsilon_{i,h}^{r,a} = 0$), where the level and the dispersion of return rates do not respond to tax reforms. The red lines display the distributional effects under a medium amount of scale dependence ($\varepsilon_{i,h}^{r,a} = 0.2$). In Appendix B.2, I also show the distributional impact when return rates react more sensitively ($\varepsilon_{i,h}^{r,a} = 0.4$).

This quantitative exercise reveals three findings. Firstly, increasing the capital gains tax for the rich induces quantitatively larger adjustments in the pre-tax return rate distribution than tax cuts for the poor. Secondly, the response of pre-tax returns to capital tax reforms can be significant. Thus, in the presence of scale dependence, capital taxation shapes the distribution of pre-tax returns, and a government trades off novel equity gains that arise from a reduction in pre-tax return inequality and novel efficiency costs (drop in the level of pre-tax returns). These

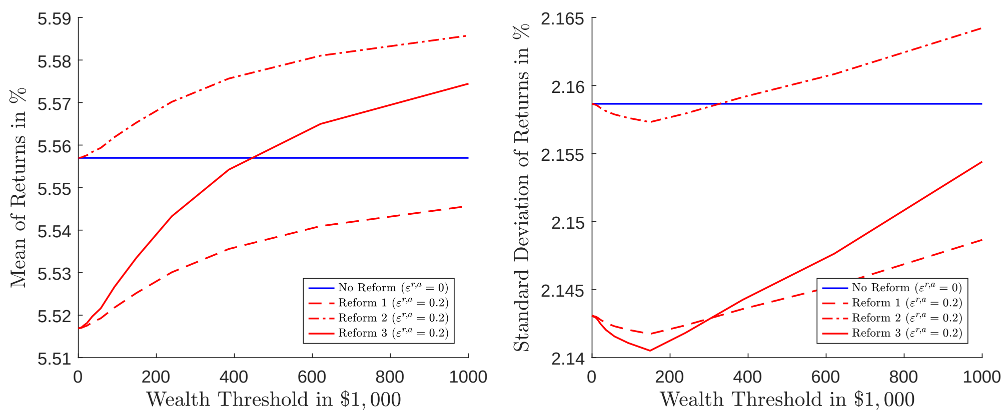

Figure 4: Left Panel: Response of Mean Pre-Tax Returns; Right Panel: Response of Standard Deviation of Pre-Tax Return

distributional responses to tax reforms may provide a source for empirically identifying the amount of scale dependence in an economy. Thirdly, the magnitude of distributional responses is larger for the level than the dispersion of pre-tax return rates. This observation indicates that efficiency margins are quantitatively more important than equity considerations. Therefore, I focus on the revenue effects of the tax reforms in the following.

Revenue effects. In Appendix B.2, I compare the different tax reforms' revenue effects to each other. Tax hikes (Reform 1) raise revenues, while tax cuts (Reform 2) reduce them. Thus, the mechanical revenue gains from increasing the current US capital gains tax dominate the negative behavioral effects. Therefore, the current US tax code is too low from a Rawlsian perspective and absent of other behavioral responses, such as tax evasion and avoidance. Moreover, the incidence of Reform 3 resembles the one of Reform 2. The reason is that Reform 3 is a mixture of the other two reforms and the effects of Reform 2 are quantitatively negligible. Based on these observations, I focus, in the following, on the reforms of type 1 that raise the capital gains tax above a certain wealth level.

The Right Panel of Figure 5 shows the substantial revenue gains from the tax hike induced by Reform 1. However, in the presence of scale dependence, this expansion of tax revenues is considerably lower. The intuition is that the tax reform lowers household's incentives to save. In the presence of scale dependence, as households save less, their return rates decline. The higher the amount of scale dependence, the larger this response (see Appendix B.2). Consequently, in the presence of scale dependence, the adjustment of return rates dampens the gains in tax revenues,

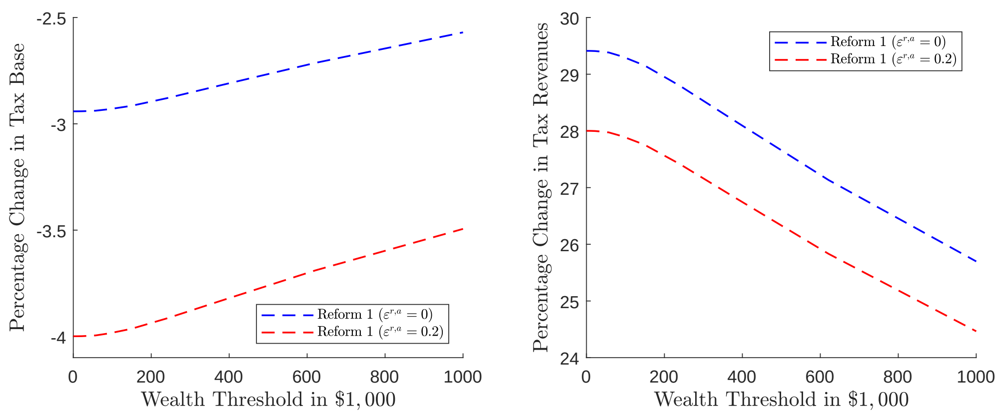

Figure 5: Left Panel: Response of Realized Capital Gains to Reform 1; Right Panel: Response of Capital Gains Tax Revenues to Reform 1

which the government realizes when raising capital taxes. Thus, the tax base shrinks more with than without scale dependence (Left Panel). This contraction is quantitatively sizable.

In this section, I have calibrated the statistical model of Section 3 to the US economy. Then, I have studied the aggregate incidence of a large set of tax reforms. Altogether, the current US capital gains tax appears too low. The quantitative analysis reveals substantial revenue gains from raising the capital gains tax rate. However, in the presence of scale dependence, these gains are considerably smaller.

5 Conclusion

This paper introduces the endogenous formation of return inequality into the optimal taxation of capital. As a microfoundation, I consider a Grossman and Stiglitz (1980) economy, in which the rich obtain higher rates of return than the poor because they are rich. Since they invest more in the financial market, wealthy households purchase more private information about the stochastic fundamentals that drive stocks' payoffs. As a result, they realize higher average rates of return. In other words, portfolio returns become scale-dependent. Although this channel may considerably raise wealth inequality, scale dependence is either neutral or provides a rationale for lower capital taxes. The reason is that scale dependence makes capital more elastic to tax reforms. I show how to adjust standard sufficient statistics that determine the elasticity of capital for scale dependence. These need to account for inequality multiplier effects between wealth and its return. Therefore, estimating the magnitude of scale dependence relative to type dependence is an important avenue for future research.

Aside from limited access to financial knowledge or advisory, other channels such as housing, liquidity constraints, insurance motives, increasing returns to wealth management, and a stock market participation cost also explain scale dependence. Due to the generality of the tax analysis, the conclusions regarding redistribution extend to these other scale dependence sources. It only depends on the magnitude of empirically observable sufficient statistics that describe the responsiveness of return rates to wealth. For pure efficiency considerations, it substantially matters to which degree the portfolios' size explains return inequality relative to type dependence. Equity considerations may even strengthen the result that taxes should be lower in the presence of scale dependence because heterogeneous returns reflect a fair reward for effort (e.g., information acquisition) and not the fortune of an inherent exceptional investment talent as with type dependence. In reality, both phenomena co-occur, but, as this paper shows, the government should address them differently.

References

Agersnap, O., and Zidar, O. M. (2020). "The tax elasticity of capital gains and revenue-maximizing rates." *National Bureau of Economic Research*.

Aiyagari, S. R. (1994). "Uninsured idiosyncratic risk and aggregate saving." *The Quarterly Journal of Economics*, *109*(3), 659–684.

Ales, L., Kurnaz, M., and Sleet, C. (2015). "Technical change, wage inequality, and taxes." *American Economic Review*, *105*(10), 3061–3101.

Arrow, K. J. (1987). "The demand for information and the distribution of income." *Probability in the Engineering and Informational Sciences*, *1*(1), 3–13.

Atkinson, A. B., and Stiglitz, J. E. (1976). "The design of tax structure: direct versus indirect taxation." *Journal of Public Economics*, *6*(1-2), 55–75.

Bach, L., Calvet, L. E., and Sodini, P. (2020). "Rich pickings? risk, return, and skill in household wealth." *American Economic Review*, *110*(9), 2703–2747.

Bakija, J. M., and Gentry, W. M. (2014). "Capital gains taxes and realizations: Evidence from a long panel of state-level data." *Working Paper*.

Benhabib, J., Bisin, A., and Zhu, S. (2011). "The distribution of wealth and fiscal policy in economies with finitely lived agents." *Econometrica*, *79*(1), 123–157.

Brülhart, M., Gruber, J., Krapf, M., and Schmidheiny, K. (2016). "Taxing wealth: Evidence from switzerland." *National Bureau of Economic Research*.

Chamley, C. (1986). "Optimal taxation of capital income in general equilibrium with infinite lives." *Econometrica*, 607–622.

Chetty, R. (2009). "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annu. Rev. Econ.*, *1*(1), 451–488.

Diamond, P. A. (1975). "A many-person ramsey tax rule." *Journal of Public Economics*, *4*(4), 335–342.

Diamond, P. A. (1998). "Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates." *American Economic Review*, 83–95.

Dowd, T., McClelland, R., and Muthitacharoen, A. (2015). "New evidence on the tax elasticity of capital gains." *National Tax Journal*, *68*(3), 511.

Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2020). "Heterogeneity and persistence in returns to wealth." *Econometrica*, *88*(1), 115–170.

Farhi, E., and Werning, I. (2010). "Progressive estate taxation." *The Quarterly Journal of Economics*, *125*(2), 635–673.

Feldstein, M., Slemrod, J., and Yitzhaki, S. (1980). "The effects of taxation on the selling of corporate stock and the realization of capital gains." *The Quarterly Journal of Economics*, *94*(4), 777–791.

Findeisen, S., and Sachs, D. (2016). "Education and optimal dynamic taxation: The role of income-contingent student loans." *Journal of Public Economics*, *138*, 1–21.

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2016). "The dynamics of inequality." *Econometrica*, *84*(6), 2071–2111.

Gerritsen, A., Jacobs, B., Rusu, A. V., and Spiritus, K. (2019). "Optimal taxation of capital income with heterogeneous rates of return." *Working Paper*.

Golosov, M., Tsyvinski, A., Werning, I., Diamond, P., and Judd, K. L. (2006). "New dynamic public finance: A user's guide [with comments and discussion]." *NBER macroeconomics annual*, *21*, 317–387.

Golosov, M., Tsyvinski, A., and Werquin, N. (2014). "A variational approach to the analysis of tax systems." Tech. rep., National Bureau of Economic Research.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). "Investment, capacity utilization, and the real business cycle." *American Economic Review*, 402–417.

Grossman, S. J., and Stiglitz, J. E. (1980). "On the impossibility of informationally efficient markets." *American Economic Review*, *70*(3), 393–408.

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo-Diaz, S., and Chen, D. (2019). "Use it or lose it: Efficiency gains from wealth taxation." *National Bureau of Economic Research*.

Hendren, N. (2016). "The policy elasticity." *Tax Policy and the Economy*, *30*(1), 51–89.

Internal Revenue Service (2020). "Soi tax stats - private foundations harmonized microdata files (ascii)." `https://www.irs.gov/statistics/soi-tax-stats-private-foundations-harmonized-microdata-files-ascii`, accessed: 2020-10-17.

Jakobsen, K., Jakobsen, K., Kleven, H., and Zucman, G. (2020). "Wealth taxation and wealth accumulation: Theory and evidence from denmark." *The Quarterly Journal of Economics*, *135*(1), 329–388.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). "The rate of return on everything, 1870–2015." *The Quarterly Journal of Economics*, *134*(3), 1225–1298.

Judd, K. L. (1985). "Redistributive taxation in a simple perfect foresight model." *Journal of Public Economics*, *28*(1), 59–83.

Kacperczyk, M., Nosal, J., and Stevens, L. (2019). "Investor sophistication and capital income inequality." *Journal of Monetary Economics*, *107*, 18–31.

Kleven, H. J. (2016). "Bunching." *Annual Review of Economics*, *8*, 435–464.

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford publications.

Krueger, D., and Ludwig, A. (2013). "Optimal progressive labor income taxation and education subsidies when education decisions and intergenerational transfers are endogenous." *American Economic Review*, *103*(3), 496–501.

Kuhn, M., Schularick, M., and Steins, U. I. (2020). "Income and wealth inequality in america, 1949–2016." *Journal of Political Economy*, *128*(9), 3469–3519.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2017). "Optimal financial knowledge and wealth inequality." *Journal of Political Economy*, *125*(2), 431–477.

Markowitz, H. (1952). "The utility of wealth." *Journal of Political Economy*, *60*(2), 151–158.

Markowitz, H. (1959). "Portfolio selection." *Investment under Uncertainty.*

Mirrlees, J. A. (1971). "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, *38*(2), 175–208.

Peress, J. (2004). "Wealth, information acquisition, and portfolio choice." *The Review of Financial Studies*, *17*(3), 879–914.

Piketty, T. (1997). "La redistribution fiscale face au chômage." *Revue française d'économie*, *12*(1), 157–201.

Piketty, T. (2014). "Capital in the 21st century."

Piketty, T., and Saez, E. (2013). "A theory of optimal inheritance taxation." *Econometrica*, *81*(5), 1851–1886.

Piketty, T., Saez, E., and Stantcheva, S. (2014). "Optimal taxation of top labor incomes: A tale of three elasticities." *American economic journal: economic policy*, *6*(1), 230–71.

Rothschild, C., and Scheuer, F. (2013). "Redistributive taxation in the roy model." *The Quarterly Journal of Economics*, *128*(2), 623–668.

Rothschild, C., and Scheuer, F. (2016). "Optimal taxation with rent-seeking." *The Review of Economic Studies*, *83*(3), 1225–1262.

Sachs, D., Tsyvinski, A., and Werquin, N. (2020). "Nonlinear tax incidence and optimal taxation in general equilibrium." *Econometrica*, *88*(2), 469–493.

Saez, E. (2001). "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, *68*(1), 205–229.

Saez, E. (2002). "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes." *Journal of Public Economics*, *83*(2), 217–230.

Saez, E., and Stantcheva, S. (2016). "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review*, *106*(1), 24–45.

Saez, E., and Stantcheva, S. (2018). "A simpler theory of optimal capital taxation." *Journal of Public Economics*, *162*, 120–142.

Saez, E., and Zucman, G. (2016). "Wealth inequality in the united states since 1913: Evidence from capitalized income tax data." *The Quarterly Journal of Economics*, *131*(2), 519–578.

Saez, E., and Zucman, G. (2019). *The triumph of injustice: How the rich dodge taxes and how to make them pay*. WW Norton & Company.

Samuelson, P. A. (1948). "Foundations of economic analysis."

Scheuer, F., and Werning, I. (2017). "The taxation of superstars." *The Quarterly Journal of Economics*, *132*(1), 211–270.

Seim, D. (2017). "Behavioral responses to wealth taxes: Evidence from sweden." *American Economic Journal: Economic Policy*, *9*(4), 395–421.

Shourideh, A. (2012). "Optimal taxation of wealthy individuals." *Working Paper*.

Stantcheva, S. (2017). "Optimal taxation and human capital policies over the life cycle." *Journal of Political Economy*, *125*(6), 1931–1990.

Stiglitz, J. E. (1982). "Self-selection and pareto efficient taxation." *Journal of Public Economics*, *17*(2), 213–240.

Verrecchia, R. E. (1982). "Information acquisition in a noisy rational expectations economy." *Econometrica*, 1415–1430.

Werning, I. (2007). "Pareto efficient income taxation." *mimeo, MIT*.

Yitzhaki, S. (1987). "The relation between return and income." *The Quarterly Journal of Economics*, *102*(1), 77–95.

Zemyan, S. M. (2012). *The classical theory of integral equations: a concise treatment*. Springer Science & Business Media.

Zoutman, F. T. (2015). "The effect of capital taxation on households' portfolio composition and intertemporal choice." *Working Paper*.

A Proofs of Section 2

A.1 Part (a) of Proposition 1

With and without scale dependence, the government solves $\underset{\tau_K}{max} \int_i \mathit{\Gamma}_i U(\tau_K; w_i)\, di$ subject to $\int_i \tau_K a_{R,i} di \geq \overline{E}$. Assume that the optimization problem is concave. Taking the derivative of the Lagrangian function $\mathcal{L} = \int_i \mathit{\Gamma}_i U(\tau_K; w_i)\, di + \lambda \left[\int_i \tau_K a_{R,i} di - \overline{E}\right]$ with respect to τ_K, the first-order condition reads as

$$\int_i (\mathit{\Gamma}_i/\lambda) \frac{dU(\tau_K; w_i)}{d\tau_K} di + \int_i a_{R,i} di = \frac{\tau_K}{1-\tau_K} \int_i a_{R,i} \zeta_i^{a_R,(1-\tau_K)} di. \tag{9}$$

With a utility function that is quasilinear in the consumption of final wealth, the first-order effects on household utility is given by $\frac{dU(\tau_K;w_i)}{d\tau_K} = -a_{R,i}$ and the shadow value of public funds λ is equal to $\int_i \mathit{\Gamma}_i di = 1$. Simplify (9) to obtain the Ramsey formula for the optimal capital gains tax.

A.2 Part (b) of Proposition 1

Without scale dependence, the average elasticity of capital income simplifies to

$$\overline{\zeta}^{a_R,(1-\tau_K)}|_{r_i} = \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} \tilde{\zeta}_i^{a_R,(1-\tau_K)} di = \tilde{\zeta}_i^{a_R,(1-\tau_K)} = \tilde{\zeta}_i^{a,(1-\tau_K)} \tag{10}$$

for constant elasticities. Define $\phi_i \equiv \frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}$ and $\Phi_i \equiv (1+\varepsilon_i^{r,a})\,\phi_i$.With scale dependence, the household elasticity of savings

$$\begin{aligned}
\zeta_i^{a,(1-\tau_K)} &\equiv \frac{dlog(a_i)}{dlog(1-\tau_K)} = \frac{dlog(a_i)}{dlog(1-\tau_K)}|_{r_i} + \frac{dlog(a_i)}{dlog(r_i)} \frac{dlog[r_i(a_i)]}{dlog(a_i)} \frac{dlog(a_i)}{dlog(1-\tau_K)} \\
&= \tilde{\zeta}_i^{a,(1-\tau_K)} + \zeta_i^{a,r}\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)} = \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}
\end{aligned}$$

and the capital income elasticity

$$\begin{aligned}
\zeta_i^{a_R,(1-\tau_K)} &\equiv \frac{dlog[a_i r_i(a_i)]}{dlog(1-\tau_K)} = \frac{dlog(a_i)}{dlog(1-\tau_K)} + \frac{dlog[r_i(a_i)]}{dlog(a_i)} \frac{dlog(a_i)}{dlog(1-\tau_K)} \\
&= (1+\varepsilon_i^{r,a})\,\zeta_i^{a,(1-\tau_K)} = (1+\varepsilon_i^{r,a})\,\phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}
\end{aligned}$$

both account for the endogenous return rate. Then, the average capital income elasticity with scale dependence

$$\overline{\zeta}^{a_R,(1-\tau_K)} = \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} (1+\varepsilon_i^{r,a})\,\phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} di = \frac{1+\varepsilon_i^{r,a}}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}} \tilde{\zeta}_i^{a,(1-\tau_K)} \tag{11}$$

is larger than the one without $\overline{\zeta}^{a_R,(1-\tau_K)} > \overline{\zeta}^{a_R,(1-\tau_K)}|_{r_i}$ for $\varepsilon_i^{r,a} > 0$.

A.3 Part (c) of Proposition 1

The response of the inequality measure $I(\tau_K)$ can be written as

$$I'(\tau_K) = -\frac{1}{1-\tau_K}\frac{\int_i a_{R,i}di \cdot \int_i (1-\Gamma_i)\, a_{R,i}\zeta_i^{a_R,(1-\tau_K)}di - \int_i \zeta_i^{a_R,(1-\tau_K)} a_{R,i}di \cdot \int_i (1-\Gamma_i)\, a_{R,i}di}{\left(\int_i a_{R,i}di\right)^2}.$$

For constant elasticities $\tilde{\zeta}_i^{a,(1-\tau_K)}$, $\zeta_i^{a,r}$, and $\varepsilon_i^{r,a}$, the capital income elasticity, $\zeta_i^{a_R,(1-\tau_K)}$, is also uniform across the population. Accordingly, the denominator of $I'(\tau_K)$ is equal to zero.

A.4 Part (d) of Proposition 1

The change in mean returns, $\mathbb{E}(r_i) = \int_i r_i(a_i)\, di$, from a tax reform $d\tau_K$ can be expressed as

$$\begin{aligned} d\mathbb{E}(r_i) &= -\int_i r_i(a_i)\frac{dlog\left[r_i(a_i)\right]}{dlog(a_i)}\frac{dlog(a_i)}{dlog(1-\tau_K)}di \cdot \frac{d\tau_K}{1-\tau_K} \\ &= -\mathbb{E}(r_i)\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K}. \end{aligned}$$

Similarly, differentiate the variance of returns, $\mathbb{V}(r_i) = \mathbb{E}(r_i^2) - \mathbb{E}(r_i)^2$,

$$\begin{aligned} d\mathbb{V}(r_i) &= -2\mathbb{E}\left(r_i^2\right)\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K} + 2\mathbb{E}(r_i)^2\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K} \\ &= -2\mathbb{V}(r_i)\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K}. \end{aligned}$$

Whenever $\zeta_i^{r,a} > 0$, $d\mathbb{E}(r_i) < 0$ and $d\mathbb{V}(r_i) < 0$.

A.5 Part (e) of Proposition 1

Optimal taxation in general equilibrium. As in A.1, one calculates the social planner's first-order condition

$$\int_i (\Gamma_i/\lambda)\frac{dU(\tau_K;w_i)}{d\tau_K}di + \int_i (\Gamma_i/\lambda)\frac{dU(\tau_K;w_i)}{dr_i}\int_{i'}\frac{dr_i}{da_{i'}}\frac{da_{i'}}{d\tau_K}di'di + \int_i a_{R,i}di = \frac{\tau_K}{1-\tau_K}\int_i a_{R,i}\zeta_i^{a_R,(1-\tau_K)}di, \tag{12}$$

where the second term on the left-hand side of (12) collects cross-effects in each households' return rates. Note that by the quasilinearity of the utility function $\frac{dU(\tau_K;w_i)}{dr_i} = (1-\tau_K)\, a_i$. Using the definition of cross-return elasticities, the first-order inter-household externalities simplify to

$$\int_i (\Gamma_i/\lambda) \frac{dU(\tau_K; w_i)}{dr_i} \int_{i'} \frac{dr_i}{da_{i'}} \frac{da_{i'}}{d\tau_K} di' di = -\int_i \Gamma_i a_{R,i} \int_{i'} \gamma_{i,i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' di,$$

leading to the optimal capital gains tax in general equilibrium.

Elasticities in general equilibrium. Observe that, aside from collecting general equilibrium externalities, one needs to adjust the elasticities. With multiplicatively separable cross-return elasticities $\gamma_{i,i'}^{r,a} = \frac{1}{r_i}\delta_{i'}^{r,a}$, the savings elasticity is

$$\begin{aligned}\zeta_i^{a,(1-\tau_K)} &= \tilde{\zeta}_i^{a,(1-\tau_K)} + \zeta_i^{a,r} \varepsilon_i^{r,a} \zeta_i^{a,(1-\tau_K)} + \int_{i'} \frac{dlog(a_i)}{dlog(r_i)} \frac{dlog[r_i(\cdot)]}{dlog(a_{i'})} \frac{dlog(a_{i'})}{dlog(1-\tau_K)} di' \\ &= \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} + \phi_i \zeta_i^{a,r} \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'.\end{aligned}$$

Multiply the left-hand side by δ_i and integrate out to get

$$\begin{aligned}\int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' &= \int_{i'} \delta_{i'}^{r,a} di' \cdot \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} + \phi_i \zeta_i^{a,r} \int_{i'} \delta_{i'}^{r,a} \frac{1}{r_{i'}} di' \cdot \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' \\ &= \int_{i'} \delta_{i'}^{r,a} di' \cdot \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)},\end{aligned}$$

where the second equality follows by the simplifying assumption that cross-effects average out $\int_{i'} \gamma_{i',i'}^{r,a} di' = 0$. Moreover, if $\delta_{i'}^{r,a}$ decreases in i' (whereas return rates increase in i'),

$$\underbrace{\mathbb{COV}\left(\frac{1}{r_{i'}}, \delta_{i'}^{r,a}\right)}_{>0} = \int_{i'} \gamma_{i',i'}^{r,a} di' - \mathbb{E}\left(\frac{1}{r_{i'}}\right) \mathbb{E}\left(\delta_{i'}^{r,a}\right) = \underbrace{-\mathbb{E}\left(\frac{1}{r_{i'}}\right)}_{<0} \mathbb{E}\left(\delta_{i'}^{r,a}\right).$$

Then, $\mathbb{E}(\delta_{i'}^{r,a}) = \int_{i'} \delta_{i'}^{r,a} di'$ must be negative and the elasticity is smaller in general than in partial equilibrium

$$\zeta_i^{a,(1-\tau_K)} = \left(1 + \zeta_i^{a,r} \phi_i \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} di'\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} < \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}. \tag{13}$$

Notice that the savings elasticity is increasing in i.

The elasticity of capital income can be written as

$$\begin{aligned}\zeta_i^{a_R,(1-\tau_K)} &= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} + \int_{i'} \frac{dlog(a_i r_i)}{dlog(r_i)} \frac{dlog[r_i(\cdot)]}{dlog(a_{i'})} \frac{dlog(a_{i'})}{dlog(1-\tau_K)} di' \\ &= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} + \left(1 + \zeta_i^{a,r}\right) \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'.\end{aligned} \tag{14}$$

Assuming positive savings elasticities, the second term on the right-hand side is, again, negative since

$$\int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' = \underbrace{\mathbb{COV}\left(\delta_{i'}^{r,a}, \zeta_{i'}^{a,(1-\tau_K)}\right)}_{<0} + \underbrace{\mathbb{E}\left(\delta_{i'}^{r,a}\right)}_{<0} \underbrace{\mathbb{E}\left(\zeta_{i'}^{a,(1-\tau_K)}\right)}_{>0} < 0.$$

Thus, in general equilibrium, one needs to downward adjust the capital income elasticity

$$\zeta_i^{a_R,(1-\tau_K)} < \left(1+\varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} < \left(1+\varepsilon_i^{r,a}\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}.$$

Furthermore, the general equilibrium welfare externalities

$$\gamma_i^{r,(1-\tau_W)} = \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'$$

are negative because $\int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' < 0$.

Comparative statics. Firstly, express the capital gains elasticity in Equation (14) as

$$\zeta_i^{a_R,(1-\tau_K)} = \underbrace{\left(1+\varepsilon_i^{r,a}\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}}_{c_1} + \underbrace{\left(1+\varepsilon_i^{r,a} + \zeta_i^{a,r} \phi_i\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} \int_{i'} \delta_{i'}^{r,a} di'}_{c_2} \cdot \frac{1}{r_i}, \tag{15}$$

where $c_1 > 0$ and $c_2 < 0$ are constants. Use this expression to write the measure of inequality that serves as a sufficient statistic for the optimal capital income tax as

$$I'\left(\tau_K^{GE}\right) = \frac{-c_2}{1-\tau_K^{GE}} \frac{\mathbb{E}\left(a_i\right) \mathbb{E}\left(\mathit{\Gamma}_i a_{R,i}\right) - \mathbb{E}\left(a_{R,i}\right) \mathbb{E}\left(\mathit{\Gamma}_i a_i\right)}{\left[\mathbb{E}\left(a_{R,i}\right)\right]^2}.$$

Notice that $\mathbb{COV}\left(\mathit{\Gamma}_i, a_{R,i}\right) < 0$, $\mathbb{COV}\left(\mathit{\Gamma}_i, a_i\right) < 0$, and, by the fact that capital income is convex in savings, $\mathbb{COV}\left(\mathit{\Gamma}_i, a_{R,i}\right) < \mathbb{COV}\left(\mathit{\Gamma}_i, a_i\right)$. Therefore, $I'\left(\tau_K^{GE}\right)$ is negative since

$$\begin{aligned}
\mathbb{E}\left(a_i\right) \mathbb{E}\left(\mathit{\Gamma}_i a_{R,i}\right) - \mathbb{E}\left(a_{R,i}\right) \mathbb{E}\left(\mathit{\Gamma}_i a_i\right) &= \mathbb{E}\left(a_i\right) \mathbb{COV}\left(\mathit{\Gamma}_i, a_{R,i}\right) - \mathbb{E}\left(a_{R,i}\right) \mathbb{COV}\left(\mathit{\Gamma}_i, a_i\right) \\
&= \underbrace{\mathbb{E}\left(a_{R,i}\right)}_{>0} \underbrace{\left[\mathbb{COV}\left(\mathit{\Gamma}_i, a_{R,i}\right) - \mathbb{COV}\left(\mathit{\Gamma}_i, a_i\right)\right]}_{<0} + \underbrace{\mathbb{E}\left(\left(1-r_i\right) a_i\right)}_{>0} \underbrace{\mathbb{COV}\left(\mathit{\Gamma}_i, a_{R,i}\right)}_{<0}
\end{aligned}$$

for $r_i \in [0,1]$.

In the following, I approximate individual and aggregate variables in general equilibrium (and evaluated at the general equilibrium tax) around the values one would obtain when having the partial equilibrium tax rate. In other words, to show that $\tau_K^{GE} > \tau_K^{PE}$, for small general equilibrium forces ($\delta_{i'}^{r,a} \approx 0$ and $\tau_K^{GE} \approx \tau_K^{PE}$), I apply a Taylor expansion to the optimal capital income tax

$$\frac{\tau_K^{GE}}{1-\tau_K^{GE}} = \frac{\mathbb{E}\left[\left(1-\mathit{\Gamma}_i\left(1+\gamma_i^{r,(1-\tau_W)}\right)\right)a_{R,i}\left(\tau_K^{GE}\right)\right]}{\mathbb{E}\left[\left(c_1+\frac{1}{r_i^{GE}}c_2\right)a_{R,i}\left(\tau_K^{GE}\right)\right]}.$$

A household's capital income in general equilibrium is approximately

$$\begin{aligned} a_{R,i}\left(\tau_K^{GE}\right) &= a_{R,i}\left(\tau_K^{PE}\right) - \left(\tau_K^{GE}-\tau_K^{PE}\right)\frac{da_{R,i}}{d\left(1-\tau_K\right)} + o\left(\tau_K^{GE}-\tau_K^{PE}\right) \\ &= a_{R,i}\left(\tau_K^{PE}\right) - \frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{GE}}\zeta_i^{a_R,(1-\tau_K)}a_{R,i}\left(\tau_K^{PE}\right) + o\left(\tau_K^{GE}-\tau_K^{PE}\right) \\ &= a_{R,i}\left(\tau_K^{PE}\right) - \frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{PE}}c_1 a_{R,i}\left(\tau_K^{PE}\right) + o\left(\tau_K^{GE}-\tau_K^{PE}\right), \end{aligned}$$

keeping in mind that the elasticities are evaluated in general equilibrium. Similarly, approximate aggregate variables

$$\mathbb{E}\left[\zeta_i^{a_R,(1-\tau_K)}a_{R,i}\left(\tau_K^{GE}\right)\right] = c_1\mathbb{E}\left[a_{R,i}\left(\tau_K^{PE}\right)\right]+c_2\mathbb{E}\left[a_i\left(\tau_K^{PE}\right)\right]-\frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{PE}}c_1^2\mathbb{E}\left[a_{R,i}\left(\tau_K^{PE}\right)\right]+o\left(\tau_K^{GE}-\tau_K^{PE}\right)$$

and

$$\begin{aligned} \mathbb{E}\left[\left(1-\mathit{\Gamma}_i\left(1+\gamma_i^{r,(1-\tau_W)}\right)\right)a_{R,i}\left(\tau_K^{GE}\right)\right] &= \mathbb{E}\left[\left(1-\mathit{\Gamma}_i\left(1+\gamma_i^{r,(1-\tau_W)}\right)\right)a_{R,i}\left(\tau_K^{PE}\right)\right] \\ &\quad - \frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{PE}}c_1\mathbb{E}\left[\left(1-\mathit{\Gamma}_i\right)a_{R,i}\left(\tau_K^{PE}\right)\right] + o\left(\tau_K^{GE}-\tau_K^{PE}\right). \end{aligned}$$

Use the fact that, in the self-confirming policy equilibrium, $\frac{\tau_K^{PE}}{1-\tau_K^{PE}} = \frac{\mathbb{E}\left[(1-\mathit{\Gamma}_i)a_{R,i}\left(\tau_K^{PE}\right)\right]}{c_1\mathbb{E}\left[a_{R,i}\left(\tau_K^{PE}\right)\right]}$ to express the general equilibrium tax in terms of the one in partial equilibrium

$$\frac{\tau_K^{GE}}{1-\tau_K^{GE}} = \frac{\tau_K^{PE}}{1-\tau_K^{PE}}\cdot\Delta + o\left(\tau_K^{GE}-\tau_K^{PE}\right), \tag{16}$$

where

$$\Delta \equiv \frac{1-\frac{\mathbb{E}\left[\mathit{\Gamma}_i\gamma_i^{r,(1-\tau_W)}a_{R,i}\left(\tau_K^{PE}\right)\right]}{\mathbb{E}\left[(1-\mathit{\Gamma}_i)a_{R,i}\left(\tau_K^{PE}\right)\right]}-\frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{PE}}c_1}{1+\frac{c_2\mathbb{E}\left[a_i\left(\tau_K^{PE}\right)\right]}{c_1\mathbb{E}\left[a_{R,i}\left(\tau_K^{PE}\right)\right]}-\frac{\tau_K^{GE}-\tau_K^{PE}}{1-\tau_K^{PE}}c_1}.$$

Noting that $\Delta > 1$ since $\gamma_i^{r,(1-\tau_W)} < 0$ and $c_2 < 0$, as defined in Equation (14), concludes the proof.

B Proofs of Section 4 and Additional Figures

B.1 Proofs of Section 4

Level of pre-tax returns. In this section, I theoretically confirm the relationship between the reform threshold and the pre-tax return distribution (mean and variance) for the three types of reforms analyzed in Section 4. Denote $\bar{i}$ as the threshold wealth percentile. For Reforms 1 and 2 ($d\tau_K > 0$ for $i \geq \bar{i}$ and $d\tau_K < 0$ for $i \leq \bar{i}$), the response of mean pre-tax returns read as

$$d\mathbb{E}\left(r_{i,h}\right) = -\mathbb{E}\left(r_{i,h} | i \geq \bar{i}\right)\left(1 - F\left(\bar{i}\right)\right)\varepsilon_{i,h}^{r,a}\zeta_{i,h}^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K}$$

and

$$d\mathbb{E}\left(r_{i,h}\right) = -\mathbb{E}\left(r_{i,h} | i \leq \bar{i}\right) F\left(\bar{i}\right)\varepsilon_{i,h}^{r,a}\zeta_{i,h}^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K},$$

respectively. The response is positive for a tax cut (Reform 2) and negative for a tax rise (Reform 1). A reduction in the capital tax raises the level of pre-tax returns, whereas an increase in capital taxes reduces them. Moreover, the absolute value of the expression decreases in $\bar{i}$. The intuition is that a higher $\bar{i}$ reduces the number of individuals treated by the reform. The overall effect of raising taxes at the top and cutting taxes at the bottom by the same amount $d\tau_K > 0$ (Tax Reform 3) is ambiguous. The response is given by

$$d\mathbb{E}\left(r_{i,h}\right) = \left[\mathbb{E}\left(r_{i,h}\right) F\left(\bar{i}\right) - \mathbb{E}\left(r_{i,h} | i > \bar{i}\right)\right]\varepsilon_{i,h}^{r,a}\zeta_{i,h}^{a,(1-\tau)}\frac{d\tau_K}{1-\tau_K}.$$

It is easy to show that the expression increases in $\bar{i}$, confirming solid lines' positive slopes.

Dispersion of pre-tax returns. The relationship between the variance of pre-tax returns and the reform threshold is less obvious. To demonstrate the nonlinearity between the distributional response and the threshold $\bar{i}$, I write the reaction of the pre-tax return variance to Reform 1 as

$$d\mathbb{V}\left(r_{i,h}\right) = -2\left[\mathbb{V}\left(r_{i,h} | i > \bar{i}\right) - \mathbb{E}\left(r_{i,h} | i \leq \bar{i}\right)\mathbb{E}\left(r_{i,h} | i > \bar{i}\right)\right]\left(1 - F\left(\bar{i}\right)\right)\varepsilon_{i,h}^{r,a}\zeta_{i,h}^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K}$$

and to Reform 2 as

$$d\mathbb{V}\left(r_{i,h}\right) = -2\left[\mathbb{V}\left(r_{i,h} | i \leq \bar{i}\right) - \mathbb{E}\left(r_{i,h} | i > \bar{i}\right)\mathbb{E}\left(r_{i,h} | i \leq \bar{i}\right)\right] F\left(\bar{i}\right)\varepsilon_{i,h}^{r,a}\zeta_{i,h}^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K}$$

Similarly, the response to Reform 3 ($d\tau_K > 0$) is given by

$$d\mathbb{V}\left(r_{i,h}\right) = -\,2\left[\mathbb{V}\left(r_{i,h}|i > \bar{i}\right) F\left(\bar{i}\right) - \mathbb{V}\left(r_{i,h}|i \leq \bar{i}\right)\left(1 - F\left(\bar{i}\right)\right)\right.$$
$$\left. + \mathbb{E}\left(r_{i,h}|i \leq \bar{i}\right) \mathbb{E}\left(r_{i,h}|i > \bar{i}\right)\left(2F\left(\bar{i}\right) - 1\right)\right] \varepsilon^{r,a}_{i,h} \zeta^{a,(1-\tau)}_{i,h} \frac{d\tau_K}{1 - \tau_K}.$$

Taking derivatives with respect to $\bar{i}$ and using the definitions of conditional means and variances, one can show that the above-described nonlinearities are present.

B.2 Additional Figures

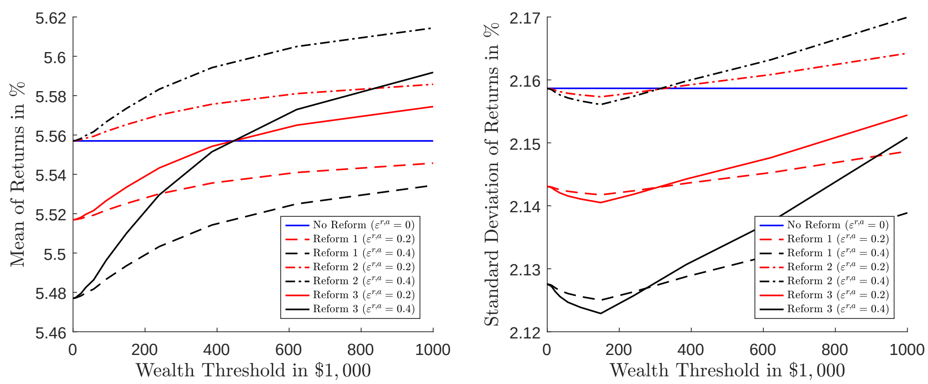

Figure 6: Left Panel: Response of Mean Pre-Tax Returns; Right Panel: Response of Standard Deviation of Pre-Tax Return

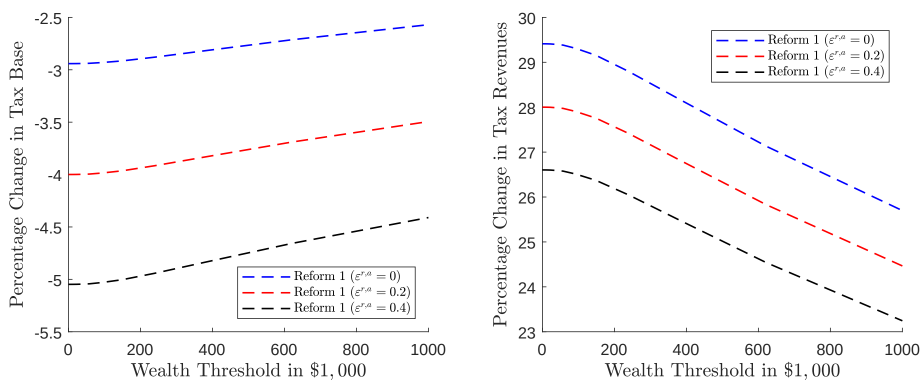

Figure 7: Left Panel: Response of Realized Capital Gains to Reform 1; Right Panel: Response of Capital Gains Tax Revenues to Reform 1

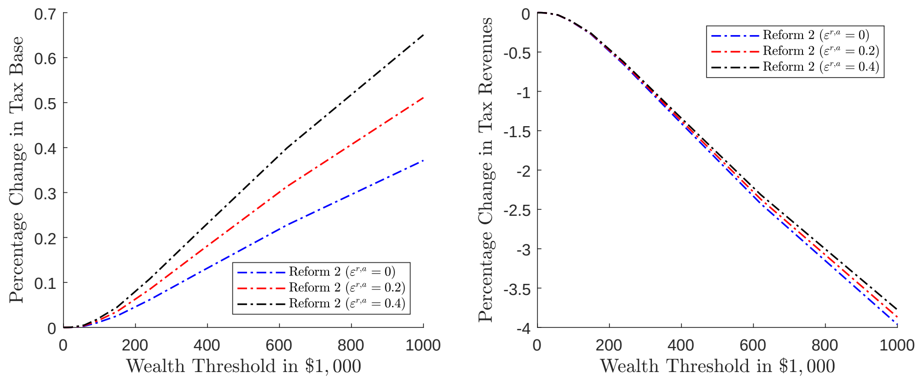

Figure 8: Left Panel: Response of Realized Capital Gains to Reform 2; Right Panel: Response of Capital Gains Tax Revenues to Reform 2

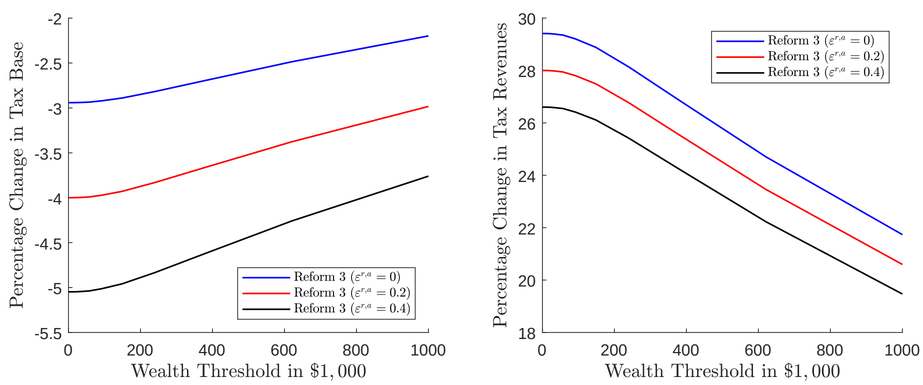

Figure 9: Left Panel: Response of Realized Capital Gains to Reform 3; Right Panel: Response of Capital Gains Tax Revenues to Reform 3

C A Dynamic Economy

In this section, I incorporate scale dependence into the dynamic bequest taxation model of Piketty and Saez (2013) that can be interpreted as a theory of capital taxation. I show that the main results from the previous section carry over. I discuss the main differences arising from a fully dynamic setting relative to the conceptual framework of Section 2. Moreover, I derive the optimal tax in general equilibrium. Finally, I deal with the role of uncertainty, which is present in the financial market of Section E.

C.1 Environment

First, I describe the economic environment closely following Piketty and Saez (2013). Consider a discrete set of periods $t \in \{0, 1, ...\}$. In each period, there lives a generation of measure one.

Preferences and technology. Each household i, t from dynasty $i \in [0, 1]$ in generation t differs in a labor skill $w_{i,t}$, which may correlate across generations. Let the distribution of skills be stationary and ergodic. Individual i, t supplies labor $l_{i,t}$ to earn a pre-tax labor income $y_{L,i,t} \equiv w_{i,t} l_{i,t}$ which is taxed linearly at rate $\tau_{L,t}$. Let E_t be an exogenous transfer. At the beginning of a period, each household receives a capital endowment (inheritance) $a_{i,t} \geq 0$ from the previous generation that carries a yield of $r_{i,t}$ and is taxed at rate, $\tau_{W,t}$.[13] Suppose the initial distribution of $a_{i,0}$ is exogenously given.

Households can take effort $x_{i,t+1}$ at a cost $v\left(x_{i,t+1}\right)$ to increase the rate of return $r'_{i,t}\left(x_{i,t}\right) > 0$ (e.g., financial advisory or financial knowledge acquisition). Let the usual monotonicity conditions hold. That is, effort choices, as well as savings, and, hence, labor and capital income are increasing the index i.[14] Intuitively, the higher an individual's hourly wage, the more she will work, and the more resources she can transfer to the retirement period. Moreover, an individual's incentives to take efforts to increase her capital gains rise with her position in the pre-tax wage distribution. Accordingly, there is scale dependence. That is, larger portfolios earn higher rates of return than smaller ones $r_{i,t} \equiv r_{i,t}\left(a_{i,t}\right)$ where $r'_{i,t}\left(a_{i,t}\right) > 0$ and $r''_{i,t}\left(a_{i,t}\right) < 0$. When the rate of return is deductible from the tax base, define $r_{i,t} \equiv r_{i,t}\left(x_{i,t}\right) - v\left(x_{i,t}\right)/a_{i,t}$. In Section E, I microfound this setup: There, returns form on a financial market in general equilibrium, making returns a function of one's own and everyone else's choices, $r_{i,t}\left(a_{i,t}, \{a_{j,t}\}_{j\in[0,1]}\right)$. For the moment, I shut down general equilibrium effects.

Household problem. Households optimally supply labor and use their after-tax, disposable income for consumption, $c_{i,t}$, and transfers into the next period (bequests), $a_{i,t+1}$, to maximize their utility $U_{i,t}\left(c_{i,t}, \underline{a}_{i,t+1}, l_{i,t}\right)$, where $a_{R,i,t+1} \equiv a_{i,t+1}\left(1 + r_{i,t+1}\right)$ and $\underline{a}_{i,t+1} \equiv a_{R,i,t+1}\left(1 - \tau_{W,t+1}\right)$

[13] With return heterogeneity, it has been noted that a tax on wealth is not equivalent to a tax on capital income, $\tau_{K,t+1}$. They yield different implications for efficiency (Guvenen et al. (2019)). That is, only when $r_{i,t+1} = r_{t+1}$ for all i, $a_{R,i,t+1}\left(1 - \tau_{W,t+1}\right) = a_{i,t+1}\left[1 + \left(1 - \tau_{K,t+1}\right) r_{t+1}\right]$ if and only if $\tau_{K,t+1} = \tau_{W,t+1}\frac{1+r_{t+1}}{r_{t+1}}$. In this paper, I disregard the important debate, which of the two policy instruments is more suitable in a given situation, and focus instead on the implications of endogenously formed return inequality for redistribution. Formally, with heterogeneous returns, a rise in the wealth tax by $d\tau_{W,t+1}$ also shifts the implied personal capital gains tax for any individual i upwards: $d\tau_{W,t+1} = d\tau_{K,i,t+1}\frac{r_{i,t+1}}{1+r_{i,t+1}} > 0$.

[14] In Section G, I address monotonicity more formally.

are the pre- and after-tax final wealth. Altogether, households solve

$$\max_{c_{i,t},l_{i,t},a_{i,t+1},x_{i,t+1}} U_{i,t}\left(c_{i,t}, a_{R,i,t+1}\left(1-\tau_{W,t+1}\right), l_{i,t}, x_{i,t+1}\right) \tag{17}$$

subject to their budget constraint $c_{i,t} + a_{i,t+1} = a_{R,i,t}\left(1-\tau_{W,t}\right) + w_{i,t}l_{i,t}\left(1-\tau_{L,t}\right) + E_t$. As returns result from effort choices ($x_{i,t+1}$), households take their rate of return $r_{i,t+1}$ as given, when choosing $a_{i,t+1}$. The first-order condition for the optimal level of $a_{i,t+1}$ is given by $\frac{\partial U_{i,t}(\cdot)}{\partial c_{i,t}} = \frac{\partial U_{i,t}(\cdot)}{\partial \underline{a}_{i,t+1}}\left(1-\tau_{W,t+1}\right)\left(1+r_{i,t+1}\right)$.

Denote $a_t \equiv \int_i a_{i,t}di$, $a_{R,t} \equiv \int_i a_{R,i,t}di$, $c_t \equiv \int_i c_{i,t}di$, and $y_{L,t} \equiv \int_i y_{i,t}di$ as the aggregate variables in period t. Suppose that the economy converges to a unique equilibrium with ergodic steady-state distributions of earnings and wealth that are independent from the initial endowments $a_{i,0}$.

C.2 Optimal Taxation in Partial Equilibrium

In the following, consider the optimal long-run tax policy in the steady-state equilibrium, (τ_W, τ_L, E). Again, denote $\Gamma_{i,t} \geq 0$ as the Pareto weights. The government maximizes the sum of weighted utilities

$$\max_{\tau_W,\tau_L} \int_i \Gamma_{i,t} U_{i,t}\left(a_{i,t}\left(1+r_{i,t}\right)\left(1-\tau_W\right) + w_{i,t}l_{i,t}\left(1-\tau_L\right) + E - a_{i,t+1},\right.$$
$$\left. a_{i,t+1}\left(1+r_{i,t+1}\right)\left(1-\tau_W\right), l_{i,t}\right) di \tag{18}$$

subject to the balanced period budget $\tau_W a_{R,t} + \tau_L y_{L,t} = E$ and scale dependence $r_{i,t} \equiv r_{i,t}\left(a_{i,t}\right)$. Observe that, for a given amount of E, τ_W and τ_L are directly linked to each other. For a budget neutral reform of the tax system, a change in τ_W triggers an according adjustment in τ_L and vice versa.

Elasticities. As before, denote the savings elasticity as $\zeta_{i,t}^{a,r} \equiv \frac{\partial log(a_{i,t})}{\partial log(r_{i,t})}$, the *own-return elasticity* as $\varepsilon_{i,t}^{r,a} \equiv \frac{\partial log[r_{i,t}(a_{i,t})]}{\partial log(a_{i,t})}$ and $\phi_{i,t} \equiv \frac{1}{1-\zeta_{i,t}^{a,r}\varepsilon_{i,t}^{r,a}} > 0$ as the measure of the *inequality multiplier effect*. It is useful to define another version of the own-return elasticity as $\varepsilon_{i,t}^{1+r,a} \equiv \frac{\partial log[1+r_{i,t}(a_{i,t})]}{\partial log(a_{i,t})}$.

With exogenous rates of return (type dependence), the elasticity of savings and initial wealth of household i reads as

$$\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} \equiv \frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_{E,r_{i,t}} = \frac{dlog\left[a_{i,t}\left(1+r_{i,t}\right)\right]}{dlog\left(1-\tau_W\right)}|_{E,r_{i,t}} > 0.$$

With endogenously formed returns (scale dependence), the elasticity of initial wealth before and after interest are given by

$$\zeta_{i,t}^{a,(1-\tau_W)} \equiv \frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_E = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} \equiv \frac{dlog\left[a_{i,t}\left(1+r_{i,t}\left(a_{i,t}\right)\right)\right]}{dlog\left(1-\tau_W\right)}|_E = \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)},$$

respectively. Observe that, due to the endogeneity of returns, $\zeta_{i,t}^{a_R,(1-\tau_W)} > \zeta_{i,t}^{a,(1-\tau_W)} > \tilde{\zeta}_{i,t}^{a,(1-\tau_W)}$. Moreover, define the long-run elasticity of aggregate wealth and labor income with respect to their retention rate as

$$\zeta^{a_R,(1-\tau_W)} \equiv \frac{dlog\left(a_{R,t}\right)}{dlog\left(1-\tau_W\right)}|_E$$

and

$$\zeta^{y_L,(1-\tau_L)} \equiv \frac{dlog\left(y_{L,t}\right)}{dlog\left(1-\tau_L\right)}|_E.$$

As in Hendren (2016), these policy elasticities e_W and e_L include own- and cross-price effects as they feature behavioral responses to a budget-neutral reform of both τ_W and τ_L. Observe that one can decompose $\zeta^{a_R,(1-\tau_W)} = \zeta_R^{a_R,(1-\tau_W)} + \zeta_H^{a_R,(1-\tau_W)} + \zeta_E^{a_R,(1-\tau_W)}$, where

$$\zeta_R^{a_R,(1-\tau_W)} \equiv \frac{1}{a_{R,t}}\int_i \left(1+R\right)a_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}di$$

is the elasticity of savings at the mean rate of return $R \equiv \int_i r_{i,t}\left(a_{i,t}\right)di$,

$$\zeta_H^{a_R,(1-\tau_W)} \equiv \frac{1}{a_{R,t}}\int_i \left[r_{i,t}\left(a_{i,t}\right)-R\right]a_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}di$$

captures the reaction of savings with return heterogeneity, and

$$\zeta_E^{a_R,(1-\tau_W)} \equiv \frac{1}{a_{R,t}}\int_i \left[\zeta_{i,t}^{a_R,(1-\tau_W)} - \tilde{\zeta}_{i,t}^{a,(1-\tau_W)}\right]\left[1+r_{i,t}\left(a_{i,t}\right)\right]a_{i,t}di$$

characterizes the effects from the endogeneity in returns. This decomposition nests the setting of Piketty and Saez (2013) in which $\zeta_H^{a_R,(1-\tau_W)} = 0$ and $\zeta_E^{a_R,(1-\tau_W)} = 0$. Observe that $\zeta_E^{a_R,(1-\tau_W)} > 0$ for $r'_{i,t}\left(a_{i,t}\right) > 0$. Hence, for a given distribution of wealth and returns the elasticity of wealth, $\zeta^{a_R,(1-\tau_W)}$, is larger under scale dependence (when returns form endogenously) than under type dependence (Part (*b*) of Proposition 1). Also note that, by the construction of scale dependence, Proposition 1 (*d*) applies: $\zeta_t^{\mathbb{V}(r),(1-\tau_W)} = 2\varepsilon_{i,t}^{r,a}\zeta_{i,t}^{a,(1-\tau_W)} > 0$ and $\zeta_t^{\mathbb{E}(r),(1-\tau_W)} = \varepsilon_{i,t}^{r,a}\zeta_{i,t}^{a,(1-\tau_W)} > 0$ for constant elasticities.

Distributional parameters. Denote $g_{i,t} \equiv \mathit{\Gamma}_{i,t}\frac{\partial U_{i,t}(\cdot)}{\partial c_{i,t}} / \int_{i'} \mathit{\Gamma}_{i',t}\frac{\partial U_{i',t}(\cdot)}{\partial c_{i',t}}di'$ as the social marginal welfare weight of an individual i, t in monetary units. Define the ratios

$$\overline{a}^{initial} \equiv \int_i g_{i,t} \frac{[1 + r_{i,t}(a_{i,t})]\, a_{i,t}}{a_{R,t}} di$$

and

$$\overline{a}^{final} \equiv \int_i g_{i,t} \frac{a_{i,t+1}}{a_{R,t}} di$$

as the distributional parameter of initial and final wealth before interest (received and left bequests). Similarly, define the distributional parameter of labor income $\overline{y}_L \equiv \int_i g_{i,t} \frac{y_{L,i,t}}{y_{L,t}} di$. For a given unweighted population mean, a small distributional parameter indicates a strong taste for redistribution. Alternatively, fix the redistributive goal of the society. Then, a high concentration of the respective variable leads to a low value of the distributional parameter.

Steady state. To derive the optimal tax formula, one needs to find the combination of tax rates that leads to no first-order welfare gain for any budget-neutral tax reform. First, I describe the set of budget-neutral tax reforms $(d\tau_W, d\tau_L, dE)$ with $dE = 0$. Accordingly, it follows from the government budget constraint that $d\tau_W$ and $d\tau_L$ relate to each other in the following fashion

$$a_{R,t} d\tau_W \left(1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}\right) = -y_{L,t} d\tau_L \left(1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}\right). \tag{19}$$

Using the envelope theorem and imposing that the first-order change in welfare equals zero $dSWF = 0$, yields an optimality condition for the capital tax

$$\int_i g_{i,t} \left[-\left(1 + \zeta_{i,t}^{a_R,(1-\tau_W)}\right) a_{R,i,t} d\tau_W + \frac{y_{L,i,t}}{y_{L,t}} \frac{1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}}{1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}} a_{R,t} d\tau_W - \frac{a_{i,t+1}}{1-\tau_W} d\tau_W\right] di = 0. \tag{20}$$

There are three effects of a rise in the capital tax. The first one describes the negative effect on initial wealth, whereas the third term the one on final wealth. The second term is the positive effect of the reduction in the labor income tax resulting from budget neutrality. Use the definitions of aggregates and distributional parameters to rewrite Equation (20)

$$-\overline{a}^{initial}\left(1 + \hat{\zeta}^{a_R,(1-\tau_W)}\right) + \frac{1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}}{1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}} \overline{y}_L - \frac{1}{1-\tau_W} \overline{a}^{final} = 0 \tag{21}$$

where $\hat{\zeta}^{a_R,(1-\tau_W)} = \int_i \zeta_{i,t}^{a_R,(1-\tau_W)} g_{i,t} \frac{a_{R,i,t}}{a_{R,t}} di / \int_i g_{i,t} \frac{a_{R,i,t}}{a_{R,t}} di$ is the welfare-weighted average initial wealth elasticity. From these arguments, Proposition 2 directly follows.

Proposition 2 (Optimal capital tax in steady state)**.** *The optimal capital tax in the long-run* *steady-state equilibrium is*

$$\tau_W = \frac{1 - \frac{\overline{a}^{initial}}{\overline{y}_L}\left(1 - \zeta^{y_L,(1-\tau_L)}\frac{\tau_L}{1-\tau_L}\right)\left(1 + \hat{\zeta}^{a_R,(1-\tau_W)} + \frac{\overline{a}^{final}}{\overline{a}^{initial}}\right)}{1 + \zeta^{a_R,(1-\tau_W)} - \frac{\overline{a}^{initial}}{\overline{y}_L}\left(1 - \zeta^{y_L,(1-\tau_L)}\frac{\tau_L}{1-\tau_L}\right)\left(1 + \hat{\zeta}^{a_R,(1-\tau_W)}\right)} \tag{22}$$

for a given labor income tax τ_L.

Proof. Appendix D.1. □

This proposition replicates the tax formula by Piketty and Saez (2013). Hence, I obtain a version of the neutrality result (Proposition 1 (*a*)) in the previous section: The sufficient statistics that describe the optimal capital tax are the same with and without scale dependence. As already mentioned, these sufficient statistics are, however, endogenous to the process of return formation and to the capital tax.

Comparative statics. To establish Part (*c*) of Proposition 1 in this economy, I introduce (a small amount of) scale dependence into an economy without scale dependence that is otherwise observationally equivalent. I analyze the comparative static that introduces scale dependence, in the following, holding the labor supply elasticity ($\zeta^{y_L,(1-\tau_L)}$), the distribution of labor income ($\overline{y}_L$), labor taxes (τ_L), and the social marginal welfare weights ($g_{i,t}$) fixed. Let the individual wealth elasticities be uncorrelated with the marginal welfare weights such that $\hat{\zeta}^{a_R,(1-\tau_W)} = \zeta^{a_R,(1-\tau_W)}$. Moreover, I take the above-described elasticities of returns ($\varepsilon_i^{r,a}$ and $\varepsilon_i^{1+r,a}$) and savings ($\zeta_i^{a,r}$ and $\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}$) as given and omit distributional effects on the aggregate elasticity ($\zeta^{a_R,(1-\tau_W)}$) that may, for instance, arise when there is a correlation between elasticities and wealth. Of course, these simplifications neglect the endogeneity of these measures to capital taxes and the allocations that will change when introducing scale dependence. However, they allow for a tractable analysis of taxes with and without scale dependence (τ_W and $\tilde{\tau}_W$, respectively).

As described, under scale dependence, the wealth elasticity has to be upward revised (Part (*b*) of Proposition 1), providing a force for lower wealth taxes. Formally, $\zeta^{a_R,(1-\tau_W)} > \zeta^{a_R,(1-\tau_W)}|_{r_i}$ since $\zeta_E^{a_R,(1-\tau_W)} > 0$. The economic intuition for this result is the same as in Section 2. Capital gains are convex under scale dependence. This convexity makes household wealth more elastic to tax reforms. Since the optimal tax rate is inversely related to this elasticity, this channel calls for lower capital taxes. For example, when wealth is infinitely concentrated ($\frac{\overline{a}^{initial}}{\overline{y}_L} \to 0$ and $\frac{\overline{a}^{final}}{\overline{y}_L} \to 0$), the capital tax rate reduces to $\tau_W = \frac{1}{1+\zeta^{a_R,(1-\tau_W)}}$. All the distributional effects on the optimal capital tax vanish. Relative to an economy with type dependence that is otherwise observationally equivalent in its wealth and returns distribution, the presence of scale dependence raises the wealth elasticity ($\zeta^{a_R,(1-\tau_W)} > \zeta^{a_R,(1-\tau_W)}|_{r_i}$). As a result, $\tau_W < \tilde{\tau}_W$.

Nonetheless, scale dependence may raise wealth inequality relative to type dependence. A lower tax under scale dependence may decrease $\overline{a}^{final}$. This channel calls for higher taxes. In other words, the expression in Proposition 2 is not in closed form. For small policy changes ($\tau_W \approx \tilde{\tau}_W$) from introducing a small amount of scale dependence ($\varepsilon_{i,t}^{r,a} \approx 0$),[15] one can use a first-order Taylor expansion to approximate aggregate wealth

$$a_{R,t}\left(\tau_W\right) = a_{R,t}\left(\tilde{\tau}_W\right)\left[1 + \frac{\tilde{\tau}_W - \tau_W}{1 - \tilde{\tau}_W}\zeta^{a_R,(1-\tau_W)}\right] + o\left(\tau_W - \tilde{\tau}_W\right), \tag{23}$$

bearing in mind that the elasticity $\zeta^{a_R,(1-\tau_W)}$ needs to account for scale dependence. Therefore a rise in the wealth tax diminishes the aggregate wealth level in the economy. Formally, $a_{R,t}\left(\tau_W\right) > a_{R,t}\left(\tilde{\tau}_W\right)$ for $\tilde{\tau}_W > \tau_W$.

Simultaneously, the wealth inequality in the society ultimately declines in response to a rise in the capital tax

$$\overline{a}^{final}\left(\tau_W\right) = \overline{a}^{final}\left(\tilde{\tau}_W\right)\frac{1 + \frac{\tilde{\tau}_W - \tau_W}{1 - \tilde{\tau}_W}\zeta^{a,(1-\tau_W)}}{1 + \frac{\tilde{\tau}_W - \tau_W}{1 - \tilde{\tau}_W}\zeta^{a_R,(1-\tau_W)}} + o\left(\tau_W - \tilde{\tau}_W\right). \tag{24}$$

If $\tilde{\tau}_W > \tau_W$, $\overline{a}^{final}\left(\tau_W\right) < \overline{a}^{final}\left(\tilde{\tau}_W\right)$ since the elasticity of aggregate wealth is larger than the aggregate savings elasticity $\zeta^{a_R,(1-\tau_W)} > \zeta^{a,(1-\tau_W)}$. Therefore, rise in the capital tax lowers the concentration of final wealth (higher $\overline{a}^{final}$). However, when one only introduces a small amount of scale dependence, this effects disappears

$$\overline{a}^{final}\left(\tau_W\right) = \overline{a}^{final}\left(\tilde{\tau}_W\right) + o\left(\tau_W - \tilde{\tau}_W\right).$$

Interestingly, the initial (weighted) inequality is also unaffected by the tax scheme

$$\overline{a}^{initial}\left(\tau_W\right) = \overline{a}^{initial}\left(\tilde{\tau}_W\right) + o\left(\tau_W - \tilde{\tau}_W\right). \tag{25}$$

The reason is that, in this specification, the decline in aggregate wealth just offsets the rise in unweighted initial inequality when individual wealth elasticities do not correlate with marginal welfare weights ($\hat{\zeta}^{a_R,(1-\tau_W)} = \zeta^{a_R,(1-\tau_W)}$). Consequently, Proposition 1 (c) approximately holds in this economy: The wealth tax in an economy with a small amount scale dependence is lower than the one in an (in terms of $\overline{a}^{initial}$ and $\overline{a}^{final}$) observationally equivalent economy without scale

[15] In Section G, I deal with a similar comparative statics exercise without imposing any assumption on the size of policy changes.

dependence as in the former the elasticity of capital is higher.[16]

Dynamic efficiency. Suppose that the government chooses $(\tau_{W,t}, \tau_{L,t})$ to maximize

$$SWF = \sum_{t=0}^{\infty} \beta^t \int_i \Gamma_{i,t} U_{i,t} \left(a_{i,t} \left(1 + r_{i,t}\right) \left(1 - \tau_{W,t}\right) + w_{i,t} l_{i,t} \left(1 - \tau_{L,t}\right) + E_t - a_{i,t+1}, \right.$$
$$\left. a_{i,t+1} \left(1 + r_{i,t+1}\right) \left(1 - \tau_{W,t+1}\right), l_{i,t}\right) di \tag{26}$$

subject to the set of period budget constraints $\tau_{W,t} a_{R,t} + \tau_{L,t} y_{L,t} = E_t$ and scale dependence $r_{i,t} \equiv r_{i,t}\left(a_{i,t}\right)$, where $\beta \in [0,1]$ denotes the generational discount rate.

To solve for the optimal policy, consider a uniform, budget-neutral reform of the tax code at a distant future point in time, T, when all variables have converged. That is $(d\tau_{W,t}, d\tau_{L,t}) = (d\tau_W, d\tau_L)$ for all $t \geq T$. Imposing that the reform has no first-order effect on social welfare, $dSWF = 0$, one obtains a dynamic version of the optimality condition from above

$$-\overline{a}^{initial} \left(1 + (1-\beta) \sum_{t=T}^{\infty} \beta^{t-T} \zeta_t^{a_R,(1-\tau_W)}\right) + \overline{y}_L (1-\beta) \sum_{t=T}^{\infty} \beta^{t-T} \frac{1 - \zeta_t^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}}{1 - \zeta_t^{y_L,(1-\tau_W)} \frac{\tau_L}{1-\tau_L}} - \frac{1}{1-\tau_W} \frac{1}{\beta} \overline{a}^{final} = 0. \tag{27}$$

Hence, in the optimal dynamic tax formula, the steady-state elasticity is now replaced with discounted elasticities. All the intuitions from the steady-state economy carry over.

C.3 Optimal Taxation in General Equilibrium

Reconsider the steady-state economy from before. Now, assume that returns are formed in general equilibrium. That is, $r_{i,t}\left(a_{i,t}, \{a_{i',t}\}_{i' \in [0,1]}\right)$. As in Section 2, define the *cross-return elasticity* as $\gamma_{i,i',t}^{r,a} \equiv \frac{\partial log(r_{i,t})}{\partial log\left(a_{i',t}\right)}$.[17] Let the cross-elasticity be multiplicatively separable $\gamma_{i,i',t}^{r,a} = \frac{1}{r_{i,t}} \delta_{i',t}^{r,a}$ (similar to the CES example of Sachs et al. (2020)). That is, a change in the savings by a household i' leads to the same change the returns of any other household i in the percentage points. In the financial market setting of Section E, this assumption holds when the costs of information acquisition are linear and all households acquire financial information. It is useful to also define another version

[16]By similar techniques, one may analyze the impact of a small change in the amount of scale dependence.

[17]Less heuristically, one may define the cross-return elasticity as the Gateaux derivative of the return functional $r_i\left(a_i, \{a_j\}_{j \in [0,1]}\right)$. That is, perturb $\{a_j\}_{j \in [0,1]}$ by the Dirac measure at i', $\delta_{i'}$,

$$\gamma_{i,i'}^{r,a} \equiv \lim_{\mu \to 0} \frac{d}{d\mu} r_i \left(a_i, \{a_j\}_{j \in [0,1]} + \mu \delta_{i'}\right).$$

The formulation of the return functional $r_i(\cdot)$ is such that there are no discontinuous jumps of $\gamma_{i,i'}^{r,a}$ at $i' = i$. Any non-infinitesimal effect of a_i on the return functional is collected in the first argument of $r_i(\cdot)$.

of the cross-return elasticity $\gamma_{i,i',t}^{1+r,a} \equiv \frac{\partial log(1+r_{i,t})}{\partial log\left(a_{i',t}\right)} = \frac{r_{i,t}}{1+r_{i,t}}\gamma_{i,i',t}^{r,a}$.

First, note that the elasticity of wealth before and after interest are augmented by general equilibrium externalities

$$\zeta_{i,t}^{a,(1-\tau_W)} = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} + \phi_{i,t}\zeta_{i,t}^{a,r}\frac{\int_{i'}\gamma_{i,i',t}^{r,a}\phi_{i',t}\tilde{\zeta}_{i',t}^{a,(1-\tau_W)}di'}{1-\int_{i'}\gamma_{i',i',t}^{r,a}\phi_{i',t}\zeta_{i',t}^{a,r}di'}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} = \left(1+\zeta_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)} + \left(1+\zeta_{i,t}^{a,1+r}\right)\int_{i'}\gamma_{i,i',t}^{1+r,a}\zeta_{i',t}^{a,(1-\tau_W)}di',$$

respectively. The aggregate and distributional variables are defined as before. The sign and the distribution of cross-return (semi-)elasticities, $\delta_{i',t}^{r,a}$, determine how the wealth elasticities are adjusted. In the model of Section E with linear information costs, $\delta_{i',t}^{r,a}$ is positive for small values of $a_{i',t}$ and negative for large ones. This situation features trickle-up forces, where a cut in the capital tax of the rich shifts economic rents from the bottom to the top.

To illustrate the implications for the wealth elasticities, assume constant elasticities $\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} = \tilde{\zeta}_{i',t}^{a,(1-\tau_W)}$, $\zeta_{i,t}^{a,r} = \zeta_{i',t}^{a,r}$, and $\varepsilon_{i,t}^{r,a} = \varepsilon_{i',t}^{r,a}$ and suppose that cross-return elasticities average out such that $\int_{i'}\gamma_{i',i',t}^{r,a}di' = 0$. Then,

$$\zeta_{i,t}^{a,(1-\tau_W)} = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}\left(1+\zeta_{i,t}^{a,r}\phi_{i,t}\frac{1}{r_{i,t}}\int_{i'}\delta_{i',t}^{r,a}di'\right) < \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} = \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)} + \frac{1+\zeta_{i,t}^{a,1+r}}{1+r_{i,t}}\int_{i'}\delta_{i',t}^{r,a}\zeta_{i',t}^{a,(1-\tau_W)}di' < \left(1+\varepsilon_{i,t}^{1+r,a}\right)\phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}.$$

Therefore, in this general equilibrium specification, wealth reacts less elastically to tax reforms relative to the partial equilibrium setting.

Taking stock of all general equilibrium externalities, the optimal tax rate is defined by the optimality condition

$$\int_i g_{i,t}\left[-\left(1+\zeta_{i,t}^{a_R,(1-\tau_W)}\right)a_{R,i,t} + \frac{y_{L,i,t}}{y_{L,t}}\frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}}a_{R,t} - \frac{a_{i,t+1}}{1-\tau_W}\left(1+\int_{i'}\gamma_{i,i',t+1}^{1+r,a}\zeta_{i',t+1}^{a,(1-\tau_W)}di'\right)\right]di = 0.$$

which can be written as

$$-\overline{a}^{initial}\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}\right) + \frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}}\overline{y}_L - \frac{1}{1-\tau_W}\overline{a}^{final}\left(1+\hat{\gamma}^{1+r,(1-\tau_W)}\right) = 0 \tag{28}$$

using the notation from above and defining $\hat{\gamma}^{1+r,(1-\tau_W)} \equiv \int_i \left(1+\zeta_{i,t}^{a,1+r}\right)\left(\int_{i'} \gamma_{i,i',t+1}^{1+r,a} \zeta_{i',t+1}^{a,(1-\tau_W)} di'\right) g_{i,t} \frac{a_{i,t+1}}{a_{R,t}} di / \int_i g_{i,t} \frac{a_{i,t+1}}{a_{R,t}} di$.

Note that $\hat{\gamma}^{1+r,(1-\tau_W)} < 0$. Thus, the general equilibrium spillovers do not only indirectly enter the cost-benefit analysis through the downward-adjusted aggregate elasticities $\zeta^{a_R,(1-\tau_W)}$ and $\hat{\zeta}^{a_R,(1-\tau_W)}$ (Proposition 1 (e)), but also directly through $\hat{\gamma}^{1+r,(1-\tau_W)}$. The latter term accounts for a first-order spillover effect on final wealth that reduces the aggregate costs of taxing wealth. This effect adds to the reduction in aggregate elasticities. To sum up, I state Proposition 3.

Proposition 3 (Optimal capital tax in general equilibrium)**.** *The optimal capital tax in the long-run steady-state general equilibrium is*

$$\tau_W^{GE} = \frac{1 - \frac{\overline{a}^{initial}}{\overline{y}_L}\left(1 - \zeta^{y_L,(1-\tau_W)} \frac{\tau_L}{1-\tau_L}\right)\left(1 + \hat{\zeta}^{a_R,(1-\tau_W)} + \frac{\overline{a}^{final}}{\overline{a}^{initial}}\left(1 + \hat{\gamma}^{1+r,(1-\tau_W)}\right)\right)}{1 + \zeta^{a_R,(1-\tau_W)} - \frac{\overline{a}^{initial}}{\overline{y}_L}\left(1 - e_L \frac{\tau_L}{1-\tau_L}\right)\left(1 + \hat{\zeta}^{a_R,(1-\tau_W)}\right)}. \tag{29}$$

for a given labor income tax τ_L*.*

Proof. Appendix D.2. □

Comparative statics. To establish the comparative statics of optimal capital taxation, as in Part (e) of Proposition 1, I follow the reasoning in Section C.2. I introduce (a small amount of) general equilibrium effects into the partial equilibrium economy with scale dependence that is otherwise observationally equivalent. I fix the labor supply elasticity ($\zeta^{y_L,(1-\tau_L)}$), the distribution of labor income ($\overline{y}_L$), labor taxes (τ_L), and the social marginal welfare weights ($g_{i,t}$). Suppose that the individual wealth elasticities do not correlate with the marginal welfare weights such that $\hat{\zeta}^{a_R,(1-\tau_W)} = \zeta^{a_R,(1-\tau_W)}$, and hold the above-described elasticities of returns ($\varepsilon_i^{r,a}$ and $\varepsilon_i^{1+r,a}$) and savings ($\zeta_i^{a,r}$ and $\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}$) constant. Moreover, I omit any distributional effects on the aggregate wealth elasticity ($\zeta^{a_R,(1-\tau_W)}$). Let the amount of scale dependence and general equilibrium forces be small ($\zeta_{i,t}^{r,a} \approx 0$ and $\delta_{i,t}^{r,a} \approx 0$).

To compare the wealth tax in partial equilibrium, τ_W^{PE}, to the one in general equilibrium, τ_W^{GE}, I approximate the endogenous distributional variables on the right-hand side of Equation (29). Again, a higher capital tax (e.g., $\tau_W^{GE} > \tau_W^{PE}$) reduces aggregate wealth (e.g., $a_{R,t}\left(\tau_W^{GE}\right) < a_{R,t}\left(\tau_W^{PE}\right)$)

$$a_{R,t}\left(\tau_W^{GE}\right) = a_{R,t}\left(\tau_W^{PE}\right)\left[1 + \frac{\tau_W^{PE} - \tau_W^{GE}}{1 - \tau_W^{PE}} \zeta^{a_R,(1-\tau_W)}\right] + o\left(\tau_W^{GE} - \tau_W^{PE}\right). \tag{30}$$

However, under the assumptions mentioned above, there are no first-order effects on initial and final wealth inequality: $\overline{a}^{initial}\left(\tau_W^{GE}\right) = \overline{a}^{initial}\left(\tau_W^{PE}\right) + o\left(\tau_W^{GE} - \tau_W^{PE}\right)$ and $\overline{a}^{final}\left(\tau_W^{GE}\right) = \overline{a}^{final}\left(\tau_W^{PE}\right) +$

$o\left(\tau_W^{GE}-\tau_W^{PE}\right)$. Accordingly, only the adjustment in the aggregate wealth elasticity, $\zeta^{a_R,(1-\tau_W)}$, and the general equilibrium externality, $\hat{\gamma}^{1+r,(1-\tau_W)}$, affect the optimal capital tax rate. To sum up, when general equilibrium forces and scale dependence are small, the optimal capital tax is higher in general equilibrium compared to the self-confirming tax in an (in terms of $\overline{a}^{initial}$ and $\overline{a}^{final}$) observationally equivalent partial equilibrium economy.[18] This result is intuitive given the presence of trickle up.

C.4 Uncertainty

In this section, I consider the Barro-Becker dynastic model extension in Piketty and Saez (2013), which allows for uncertainty in the rates of return $r_{i,t}$. In this framework, individuals do not only care about their well-being, but also about the one of their children. As before, the government chooses a linear, deterministic tax system $(\tau_{L,t},\tau_{W,t},E_t)$. Household i in period t optimally chooses $(l_{i,t},a_{i,t+1},e_{i,t})$ to maximize $U_{i,t}=u_{i,t}(c,l,e)+\beta\mathbb{E}_t[U_{i,t+1}]$, where $\beta<1$, subject to $c_{i,t}+a_{i,t+1}=(1-\tau_{W,t})\,a_{R,i,t}+(1-\tau_{L,t})\,y_{L,i,t}+E_t$. For any $a_{i,t+1}\geq 0$, the Euler equation reads as $\frac{\partial u_{i,t}(\cdot)}{\partial c_{i,t}}a_{i,t+1}=\beta(1-\tau_{W,t+1})\,\mathbb{E}_t\left[a_{R,i,t+1}\frac{\partial u_{i,t+1}}{\partial c_{i,t+1}}\right]$. In the beginning of period $t+1$, stochastic returns have realized so that one can summarize the set of Euler equations as $\overline{a}_{t+1}^{final}=\beta(1-\tau_{W,t+1})\,\overline{a}_{t+1}^{initial}$ with the definitions from the deterministic version of the model $\overline{a}_{t+1}^{initial}\equiv\int_i g_{i,0}\frac{a_{R,i,t+1}}{a_{R,t+1}}di$ and $\overline{a}_{t+1}^{final}\equiv\int_i g_{i,0}\frac{a_{i,t+1}}{a_{R,t+1}}di$ and Pareto weights $\{\mathit{\Gamma}_{0,i}\}_{i\in[0,1]}$.

Suppose that the economy features an ergodic equilibrium with long-run variables independent from initial values. Let tax policies as well as individual choices converge. In the following, I consider the utilitarian $(\mathit{\Gamma}_{0,i}=1)$ optimal long-run policy in the ergodic steady-state equilibrium. Suppose, without loss of generality, that this equilibrium is reached in period 0. The government chooses (τ_L,τ_W,E) to maximize the steady-state discounted expected social welfare

$$SWF_\infty\equiv\sum_{t=0}^{\infty}\beta^t\mathbb{E}\left[u_{i,t}\left((1-\tau_W)\,a_{R,i,t}+(1-\tau_L)\,y_{L,i,t}+E-a_{i,t+1},l_{i,t}\right)\right]\tag{31}$$

subject to $\tau_W a_{R,t}+\tau_L y_{L,t}=E$. The optimal tax system can be described by the optimality

[18] Using similar approximations, one may evaluate the impact of a small change in the amount of general equilibrium forces.

condition

$$dSWF_{\infty} = 0 = \mathbb{E}\left[\frac{\partial u_{i,0}(\cdot)}{\partial c_{i,0}}(1-\tau_W)\, da_{R,i,0}\right] - \mathbb{E}\left[\frac{\partial u_{i,0}(\cdot)}{\partial c_{i,0}} a_{R,i,0} d\tau_W\right] - \sum_{t=0}^{\infty}\beta^{t+1}\mathbb{E}\left[\frac{\partial u_{i,t+1}(\cdot)}{\partial c_{i,t+1}} a_{R,i,t} d\tau_W\right] - \sum_{t=0}^{\infty}\beta^{t}\mathbb{E}\left[\frac{\partial u_{i,t}(\cdot)}{\partial c_{i,t}} y_{L,i,t} d\tau_L\right]$$

which, using the individual's first-order conditions and budget neutrality of the tax reform and defining $\zeta_i^{a_R,(1-\tau_W)} \equiv \frac{dlog\left(a_{R,i,0}\right)}{dlog(1-\tau_W)}$, simplifies to

$$0 = -\sum_{t=0}^{\infty}\beta^{t}\mathbb{E}\left[\frac{\partial u_{i,0}(\cdot)}{\partial c_{i,0}} a_{R,i,0}\left(1+\zeta_i^{a_R,(1-\tau_W)}\right)\right] - \sum_{t=0}^{\infty}\beta^{t}\mathbb{E}\left[\frac{\partial u_{i,t}(\cdot)}{\partial c_{i,t}}\frac{a_{i,t+1}}{1-\tau_W} + \frac{\partial u_{i,t}(\cdot)}{\partial c_{i,t}} a_{R,t}\frac{\left(1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}\right)}{\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)}\frac{y_{L,i,t}}{y_{L,t}}\right] \tag{32}$$

Since the economy is in the ergodic steady state, the optimal tax formula reads as

$$\tau_W = \frac{1-\frac{(1-\beta)\overline{a}^{initial}}{\overline{y}_L}\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}+\frac{\overline{a}^{final}}{(1-\beta)\overline{a}^{initial}}\right)}{1+\zeta^{a_R,(1-\tau_W)}-\frac{(1-\beta)\overline{a}^{initial}}{\overline{y}_L}\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}\right)} \tag{33}$$

with the only difference to Proposition 2 that $\overline{a}^{initial}$ is weighted by $(1-\beta)$ to account for the fact that one discounts the costs of taxing future generations. Altogether, including uncertainty into the economy does not alter the implications of endogenous return inequality.

D Proofs of Section C

D.1 Optimal Linear Wealth Taxation in Partial Equilibrium

Elasticities. In the presence of scale dependence, the elasticity of initial wealth before and after interest can be derived as

$$\zeta_{i,t}^{a,(1-\tau_W)} = \frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_{E,r_{i,t}} + \frac{dlog\left(a_{i,t}\right)}{dlog\left(r_{i,t}\right)}\frac{dlog\left[r_{i,t}\left(a_{i,t}\right)\right]}{dlog\left(a_{i,t}\right)}\frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_E = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} \tag{34}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} = \frac{dlog\left[1+r_{i,t}\left(a_{i,t}\right)\right]}{dlog\left(1-\tau_W\right)} + \zeta_{i,t}^{a,(1-\tau_W)} = \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)}, \tag{35}$$

respectively.

Optimal capital tax in steady state. Budget neutrality of the tax reform implies

$$d\tau_W a_{R,t} + \tau_W da_{R,t} = -d\tau_L y_{L,t} - \tau_L dy_{L,t}$$
$$\iff d\tau_W a_{R,t}\left(1 - \frac{1-\tau_W}{a_{R,t}}\frac{da_{R,t}}{d\left(1-\tau_W\right)}\frac{\tau_W}{1-\tau_W}\right) = -d\tau_L y_{L,t}\left(1 - \frac{1-\tau_L}{y_{L,t}}\frac{dy_{L,t}}{d\left(1-\tau_L\right)}\frac{\tau_L}{1-\tau_L}\right),$$

which simplifies to Equation (19).

To obtain Equation (20), plug the households' first-order conditions $\frac{\partial U_{i,t}(\cdot)}{\partial c_{i,t}} = \frac{\partial U_{i,t}(\cdot)}{\partial \underline{a}_{i,t+1}} \frac{\left(1-\tau_{W,t+1}\right)a_{R,i,t+1}}{a_{i,t+1}}$ and Equation (19) into

$$dSWF = \int_i \Gamma_{i,t}\frac{\partial U_{i,t}}{\partial c_{i,t}}\left[\left(1-\tau_W\right)da_{R,i,t} - a_{R,i,t}d\tau_W - w_{i,t}l_{i,t}d\tau_L\right]di - \int_i \Gamma_{i,t}\frac{\partial U_{i,t}}{\partial \underline{a}_{i,t+1}}a_{R,i,t+1}d\tau_W di$$
$$= \int_i g_{i,t}\left[\frac{1-\tau_W}{a_{R,i,t}}\frac{da_{R,i,t}}{d\tau_W}a_{R,i,t}d\tau_W - a_{R,i,t}d\tau_W + \frac{y_{L,i,t}}{y_{L,t}}\frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_L)}\frac{\tau_L}{1-\tau_L}}a_{R,t}d\tau_W - \frac{a_{i,t+1}}{1-\tau_W}d\tau_W\right]di,$$

and set this expression equal to zero. Equation (21) follows from

$$0 = \int_i g_{i,t}\left[-\left(1+\zeta_{i,t}^{a_R,(1-\tau_W)}\right)a_{R,i,t}d\tau_W + \frac{y_{L,i,t}}{y_{L,t}}\frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_L)}\frac{\tau_L}{1-\tau_L}}a_{R,t}d\tau_W - \frac{a_{i,t+1}}{1-\tau_W}d\tau_W\right]di$$
$$= -\int_i g_{i,t}\frac{a_{R,i,t}}{a_{R,t}}di\left(1+\frac{\int_i g_{i,t}\zeta_{i,t}^{a_R,(1-\tau_W)}\frac{a_{R,i,t}}{a_{R,t}}di}{\int_i g_{i,t}\frac{a_{R,i,t}}{a_{R,t}}di}\right) + \frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_L)}\frac{\tau_L}{1-\tau_L}}\int_i g_{i,t}\frac{y_{L,i,t}}{y_{L,t}}di - \frac{1}{1-\tau_W}\int_i \frac{g_{i,t}a_{i,t+1}}{a_{R,t}}di.$$

Rearrange this equation to get the optimal wealth tax in Proposition 2.

Comparative statics. Now, I approximate individual and aggregate variables in in the presence of scale dependence (and evaluated at optimal tax rate) around the values that would emerge without scale dependence. Memorizing that the elasticities account for the presence of scale dependence, household wealth is approximately given by

$$a_{R,i,t}\left(\tau_W\right) = a_{R,i,t}\left(\tilde{\tau}_W\right) + \left(\tau_W - \tilde{\tau}_W\right)\frac{da_{R,i,t}}{d\tau_W} + o\left(\tau_W - \tilde{\tau}_W\right)$$
$$= a_{R,i,t}\left(\tilde{\tau}_W\right)\left[1 + \frac{\tilde{\tau}_W - \tau_W}{1-\tilde{\tau}_W}\zeta_{i,t}^{a_R,(1-\tau_W)}\right] + o\left(\tau_W - \tilde{\tau}_W\right).$$

Integrate out to get Equation (23)

$$a_{R,t}\left(\tau_W\right) = a_{R,t}\left(\tilde{\tau}_W\right) + \frac{\tilde{\tau}_W - \tau_W}{1-\tilde{\tau}_W}\int_i \zeta_{i,t}^{a_R,(1-\tau_W)}a_{R,i,t}\left(\tilde{\tau}_W\right)di + o\left(\tau_W - \tilde{\tau}_W\right).$$

Plug Equation (23) and

$$\int_i g_{i,t}a_{i,t}\left(\tau_W\right)di = \int_i g_{i,t}a_{i,t}\left(\tilde{\tau}_W\right)di + \frac{\tilde{\tau}_W - \tau_W}{1-\tilde{\tau}_W}\int_i g_{i,t}\zeta_{i,t}^{a,(1-\tau_W)}a_{i,t}\left(\tilde{\tau}_W\right)di + o\left(\tau_W - \tilde{\tau}_W\right)$$

into

$$\overline{a}^{final}(\tau_W) = \frac{\int_i g_{i,t} a_{i,t}(\tau_W)\,di}{a_{R,t}(\tau_W)} = \frac{a_{R,t}(\tilde{\tau}_W)}{a_{R,t}(\tau_W)}\left[\overline{a}^{final}(\tilde{\tau}_W) + \frac{\tilde{\tau}_W - \tau_W}{1-\tilde{\tau}_W}\int_i g_{i,t}\zeta_{i,t}^{a,(1-\tau_W)}\frac{a_{i,t}(\tilde{\tau}_W)}{a_{R,t}(\tilde{\tau}_W)}di\right] + o(\tau_W - \tilde{\tau}_W)$$
$$= \overline{a}^{final}(\tilde{\tau}_W)\frac{1}{1+\frac{\tilde{\tau}_W-\tau_W}{1-\tilde{\tau}_W}\zeta^{a_R,(1-\tau_W)}}\left[1+\frac{\tilde{\tau}_W-\tau_W}{1-\tilde{\tau}_W}\hat{\zeta}^{a,(1-\tau_W)}\right] + o(\tau_W-\tilde{\tau}_W),$$

where I use the definition of $\hat{\zeta}^{a,(1-\tau_W)} \equiv \int_i \zeta_{i,t}^{a,(1-\tau_W)} g_{i,t}\frac{a_{i,t}}{a_{R,t}}di / \int_i g_{i,t}\frac{a_{i,t}}{a_{R,t}}di$. Assuming that the savings elasticities are uncorrelated with the marginal welfare weights $\hat{\zeta}^{a,(1-\tau_W)} = \zeta^{a,(1-\tau_W)}$, Equation (24) follows.

Proceed along the same lines, to obtain Equation (25)

$$\overline{a}^{initial}(\tau_W) = \frac{\int_i g_{i,t} a_{R,i,t}(\tau_W)\,di}{a_{R,t}(\tau_W)} = \overline{a}^{initial}(\tilde{\tau}_W)\frac{a_{R,t}(\tilde{\tau}_W)}{a_{R,t}(\tau_W)}\left[1+\frac{\tilde{\tau}_W-\tau_W}{1-\tilde{\tau}_W}\hat{\zeta}^{a_R,(1-\tau_W)}\right] + o(\tau_W-\tilde{\tau}_W).$$

Then, $\overline{a}^{initial}(\tau_W) = \overline{a}^{initial}(\tilde{\tau}_W) + o(\tau_W - \tilde{\tau}_W)$, for $\hat{\zeta}^{a_R,(1-\tau_W)} = \zeta^{a_R,(1-\tau_W)}$.

Dynamic efficiency. Plug the households' first order conditions and Equation (19) into $dSWF = 0$ to get

$$0 = \sum_{t=T}^{\infty}\beta^t\int_i \Gamma_{i,t}\frac{\partial U_{i,t}}{\partial c_{i,t}}\left[(1-\tau_W)\,da_{R,i,t} - a_{R,i,t}d\tau_W - y_{L,i,t}d\tau_L\right]di - \sum_{t=T-1}^{\infty}\beta^t\int_i \Gamma_{i,t}\frac{\partial U_{i,t}}{\partial \underline{a}_{i,t+1}}a_{R,i,t+1}d\tau_W di$$
$$= -\sum_{t=T}^{\infty}\beta^t\int_i g_{i,t}\left[\frac{a_{R,i,t}}{a_{R,t}}\left(1+\zeta_{i,t}^{a_R,(1-\tau_W)}\right) + \frac{y_{L,i,t}}{y_L}\frac{1-\zeta_t^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta_t^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}}\right]di - \frac{1}{1-\tau_W}\sum_{t=T-1}^{\infty}\beta^t\int_i g_{i,t}\frac{a_{i,t+1}}{a_{R,t}}di$$

and use the definitions of the distributional parameters to show Equation (27).

D.2 Optimal Linear Wealth Taxation in General Equilibrium

Elasticities. The general equilibrium savings elasticity is given by

$$\zeta_{i,t}^{a,(1-\tau_W)} = \frac{dlog(a_{i,t})}{dlog(1-\tau_W)}|_{E,r_{i,t}} + \zeta_{i,t}^{a,r}\varepsilon_{i,t}^{r,a}\zeta_{i,t}^{a,(1-\tau_W)} + \zeta_{i,t}^{a,r}\int_{i'}\gamma_{i,i',t}^{r,a}\zeta_{i',t}^{a,(1-\tau_W)}di'$$
$$= \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} + \frac{1}{r_{i,t}}\phi_{i,t}\zeta_{i,t}^{a,r}\int_{i'}\delta_{i'}^{r,a}\zeta_{i',t}^{a,(1-\tau_W)}di',$$

using the multiplicatively separable cross-return elasticities $\gamma^{r,a}_{i,i',t} = \frac{1}{r_{i,t}}\delta^{r,a}_{i'}$. One can simplify the second term on the right-hand side to

$$\int_{i'} \delta^{r,a}_{i'} \zeta^{a,(1-\tau_W)}_{i',t} di' = \int_{i'} \delta^{r,a}_{i',t} \phi_{i',t} \tilde{\zeta}^{a,(1-\tau_W)}_{i',t} di' + \int_{i'} \delta^{r,a}_{i',t} \phi_{i',t} \zeta^{a,r}_{i',t} \frac{1}{r_{i',t}} di' \int_{i''} \delta^{r,a}_{i'',t} \tilde{\zeta}^{a,(1-\tau_W)}_{i'',t} di''$$
$$= \frac{1}{1 - \int_{i'} \delta^{r,a}_{i',t} \phi_{i',t} \zeta^{a,r}_{i',t} \frac{1}{r_{i',t}} di' \int_{i''}} \int_{i'} \delta^{r,a}_{i',t} \phi_{i',t} \tilde{\zeta}^{a,(1-\tau_W)}_{i',t} di'.$$

The wealth elasticity can be derived as

$$\zeta^{a_R,(1-\tau_W)}_{i,t} = \left(1 + \varepsilon^{1+r,a}_{i,t}\right) \zeta^{a,(1-\tau_W)}_{i,t} + \int_{i'} \frac{dlog\left(a_{i,t}\left(1 + r_{i,t}\left(a_{i,t}\right)\right)\right)}{dlog\left(1 + r_{i,t}\right)} \frac{dlog\left[1 + r_{i,t}\left(\cdot\right)\right]}{dlog\left(a_{i',t}\right)} \frac{dlog\left(a_{i',t}\right)}{dlog\left(1 - \tau_W\right)} |_E \, di'$$
$$= \left(1 + \varepsilon^{1+r,a}_{i,t}\right) \zeta^{a,(1-\tau_W)}_{i,t} + \left(1 + \zeta^{a,1+r}_{i,t}\right) \int_{i'} \gamma^{1+r,a}_{i,i',t} \zeta^{a,(1-\tau_W)}_{i',t} di'.$$

Under the assumption that $\tilde{\zeta}^{a,(1-\tau_W)}_{i,t}$, $\zeta^{a,r}_{i,t}$, and $\varepsilon^{r,a}_{i,t}$ are constant and cross-return elasticities average out $\int_{i'} \gamma^{r,a}_{i',i',t} di' = 0$, these expressions simplify to

$$\zeta^{a,(1-\tau_W)}_{i,t} = \phi_{i,t} \tilde{\zeta}^{a,(1-\tau_W)}_{i,t} \left(1 + \frac{1}{r_{i,t}} \phi_{i,t} \zeta^{a,r}_{i,t} \int_{i'} \delta^{r,a}_{i'} di'\right) < \phi_{i,t} \tilde{\zeta}^{a,(1-\tau_W)}_{i,t} \tag{36}$$

and

$$\zeta^{a_R,(1-\tau_W)}_{i,t} = \left(1 + \varepsilon^{1+r,a}_{i,t}\right) \zeta^{a,(1-\tau_W)}_{i,t} + \frac{1 + \zeta^{a,1+r}_{i,t}}{1 + r_{i,t}} \phi_{i,t} \tilde{\zeta}^{a,(1-\tau_W)}_{i,t} \int_{i'} \delta^{r,a}_{i',t} di' < \left(1 + \varepsilon^{1+r,a}_{i,t}\right) \phi_{i,t} \tilde{\zeta}^{a,(1-\tau_W)}_{i,t} \tag{37}$$

since $\int_{i'} \delta^{r,a}_{i'} di' < 0$ for $\int_{i'} \gamma^{r,a}_{i',i',t} di' = \underbrace{\mathbb{COV}\left(\frac{1}{r_{i',t}}, \delta^{r,a}_{i'}\right)}_{>0} + \int_{i'} \delta^{r,a}_{i'} di' \cdot \underbrace{\int_{i'} \frac{1}{r_{i',t}} di'}_{>0} = 0.$

Optimal capital tax in steady state. Observe that there are inter-household welfare externalities from the endogeneity of each household's return rate in other households' savings

$$\frac{\partial U_{i,t}}{\partial \underline{a}_{i,t+1}} a_{R,i,t+1} \int_{i'} \frac{dlog\left[a_{i,t}\left(1 + r_{i,t}\left(a_{i,t}\right)\right)\right]}{dlog\left(1 + r_{i,t}\right)} \frac{dlog\left[1 + r_{i,t}\left(\cdot\right)\right]}{dlog\left(a_{i',t}\right)} \frac{dlog\left(a_{i',t}\right)}{dlog\left(1 - \tau_W\right)} |_E \, di'$$
$$= \frac{a_{i,t+1}}{1 - \tau_W} \frac{\partial U_{i,t}\left(\cdot\right)}{\partial c_{i,t}} \left(1 + \zeta^{a,1+r}_{i,t}\right) \int_{i'} \gamma^{1+r,a}_{i,i',t} \zeta^{a,(1-\tau_W)}_{i',t} di'.$$

Insert this equation and, as before, the households' first-order conditions and Equation (19) into

$$dSWF = \int_i \mathit{\Gamma}_{i,t} \frac{\partial U_{i,t}}{\partial c_{i,t}} \left[\left(1 - \tau_W\right) da_{R,i,t} - a_{R,i,t} d\tau_W - w_{i,t} l_{i,t} d\tau_L\right] di - \int_i \mathit{\Gamma}_{i,t} \frac{\partial U_{i,t}}{\partial \underline{a}_{i,t+1}} a_{R,i,t} d\tau_W di$$
$$+ \int_i \mathit{\Gamma}_{i,t} \frac{\partial U_{i,t}}{\partial \underline{a}_{i,t+1}} a_{R,i,t+1} \int_{i'} \frac{dlog\left[a_{i,t}\left(1 + r_{i,t}\left(a_{i,t}\right)\right)\right]}{dlog\left(1 + r_{i,t}\right)} \frac{dlog\left[1 + r_{i,t}\left(\cdot\right)\right]}{dlog\left(a_{i',t}\right)} \frac{dlog\left(a_{i',t}\right)}{dlog\left(1 - \tau_W\right)} |_E \, di' di.$$

Set this expression equal to zero and use the definitions of the distributional parameters to get Equation (28). Equation (29) follows from rearranging Equation (28).

Comparative statics. As in the partial equilibrium, approximate household savings and wealth in general equilibrium as

$$a_{i,t}\left(\tau_W^{GE}\right) = a_{i,t}\left(\tau_W^{PE}\right) + \frac{\tau_W^{PE} - \tau_W^{GE}}{1 - \tau_W^{PE}} a_{i,t}\left(\tau_W^{PE}\right) \zeta_{i,t}^{a,(1-\tau_W)} + o\left(\tau_W^{GE} - \tau_W^{PE}\right)$$

and

$$a_{R,i,t}\left(\tau_W^{GE}\right) = a_{R,i,t}^{PE}\left(\tau_W^{PE}\right) + \frac{\tau_W^{PE} - \tau_W^{GE}}{1 - \tau_W^{PE}} a_{R,i,t}\left(\tau_W^{PE}\right) \zeta_{i,t}^{a_R,(1-\tau_W)} + o\left(\tau_W^{GE} - \tau_W^{PE}\right),$$

where, again, the elasticities are evaluated in general equilibrium. Integrate out the second expression to get Equation (30).

Moreover, initial wealth can be written as

$$\begin{aligned}
\overline{a}^{initial}\left(\tau^{GE}\right) &= \frac{\int_i g_{i,t} a_{R,i,t}\left(\tau^{PE}\right) di + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \int_i g_{i,t} a_{R,i,t}\left(\tau^{PE}\right) \zeta_{i,t}^{a_R,(1-\tau_W)} di}{\int_i a_{R,i,t}\left(\tau^{PE}\right) di + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \int_i a_{R,i,t}\left(\tau^{PE}\right) \zeta_{i,t}^{a_R,(1-\tau_W)} di} + o\left(\tau_W^{GE} - \tau_W^{PE}\right) \\
&= \overline{a}^{initial}\left(\tau^{PE}\right) \frac{1 + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \hat{\zeta}^{a_R,(1-\tau_W)}}{+\frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \zeta^{a_R,(1-\tau)}} + o\left(\tau_W^{GE} - \tau_W^{PE}\right).
\end{aligned}$$

Therefore, for $\hat{\zeta}^{a_R,(1-\tau_W)} = \zeta^{a_R,(1-\tau_W)}$, $\overline{a}^{initial}\left(\tau^{GE}\right) = \overline{a}^{initial}\left(\tau^{PE}\right) + o\left(\tau_W^{GE} - \tau_W^{PE}\right)$. Similarly, final wealth

$$\begin{aligned}
\overline{a}^{final}\left(\tau^{GE}\right) &= \frac{\int_i g_{i,t} a_{i,t}\left(\tau^{PE}\right) di + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \int_i g_{i,t} a_{i,t}\left(\tau^{PE}\right) \zeta_{i,t}^{a,(1-\tau_W)} di}{\int_i a_{R,i,t}\left(\tau^{PE}\right) di + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \int_i a_{R,i,t}\left(\tau^{PE}\right) \zeta_{i,t}^{a_R,(1-\tau_W)} di} + o\left(\tau_W^{GE} - \tau_W^{PE}\right) \\
&= \overline{a}^{final}\left(\tau^{PE}\right) \frac{1 + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \hat{\zeta}^{a,(1-\tau_W)}}{1 + \frac{\tau_W^{PE} - \tau_W^{GE}}{1-\tau_W^{PE}} \zeta^{a_R,(1-\tau)}} + o\left(\tau_W^{GE} - \tau_W^{PE}\right)
\end{aligned}$$

simplifies to $\overline{a}^{initial}\left(\tau^{GE}\right) = \overline{a}^{initial}\left(\tau^{PE}\right) + o\left(\tau_W^{GE} - \tau_W^{PE}\right)$ for $\varepsilon_{i,t}^{r,a} \approx 0$ and $\delta_{i,t}^{r,a} \approx 0$.

E The Financial Market

In this section, I develop a general equilibrium financial market model, which serves as a microfoundation for the endogenous formation of return inequality described earlier. Recall that households work in the first period and can transfer resources into the next period by saving parts of their labor income. In the following setting, the returns on savings form on a financial market with imperfect information. For a given amount of savings, households choose their optimal investment portfolio and can acquire information about the stochastic returns on the financial market. This setting gives rise to inequality in the returns to investment. As high-income individuals decide to save more than low-income individuals, they have an incentive to acquire more financial knowledge, which allows them to generate higher (risk-adjusted) returns.

As standard in generational models (e.g., Piketty and Saez (2013)), I subdivide the investment period into $h = 1, ..., H + 1$ subperiods. For instance, for $H = 30$, the working life has a duration of 30 years. In the following environment, this means that, during their working life, households repeatedly interact on the financial market. In particular, they can adjust their portfolio and their financial knowledge. Between subperiods, there is no time discounting.

E.1 Environment

I model the financial market in each subperiod h as in Peress (2004) version of the Grossman and Stiglitz (1980) economy. The general equilibrium model features individuals, who differ in their initial wealth, $a_{i,h}$, which depends on initial savings $a_{i,1}$ and returns realized before h, a financial market with public and private signals about stochastic returns, and endogenous inequality in investment returns. The main goal is to justify the reduced form of investment returns as a function of initial savings $r\left(a_i, \{a_j\}_{j\in[0,1]}\right)$. Whenever I drop the subperiod index h, I refer to the first subperiod ($h = 1$).

Payoff structure. In subperiod h, household $i \in [0, 1]$ invests $a_{i,h}$ on a financial market. As in Grossman and Stiglitz (1980), there are two assets: one risk-free asset (bond) and one risky asset (stock). In each subperiod h, households purchase a costly private signal about the stock's payoff and observe a public signal (price). After that, they decide on how much to invest in the risky and the risk-free asset. In this class of models, there exists no closed-form solution for the rational expectations equilibrium in settings that go beyond constant absolute risk aversion (CARA) utilities. In these models, this issue is also present when one considers redistributive taxation. Therefore, I adopt the idea by Peress (2004) who scales the economy with a parameter z. For a small z, one can approximately solve the model in closed form for arbitrary preferences and nonlinear taxes.[19]

In each subperiod, there is a risk-free asset in infinitely elastic supply that delivers a return

[19] This procedure is similar to the time increment dt in continuous-time models.

of $r_h^f z$. The risky asset has an endogenous price P_h and a random payoff Π_h that is log-normally distributed with mean $b_h z$ and variance $\sigma^2 z$, where $log\left(\Pi_h\right) \equiv \pi_h z$. The mean payoff is normally distributed $b_h \sim \mathcal{N}\left(\mathbb{E}\left(b\right), \sigma_b^2\right)$. In other words, in each subperiod, nature draws a stochastic fundamental of the economy that drives stock returns. For simplicity, I assume that the draws of b_h are uncorrelated over time.

Define $r_{i,h}^p z$ as the realized investment return of household i in subperiod h. For a small z (e.g., $z = 1/H$), $r_{i,h}^p z$ is small so that one can neglect nonlinearities as follows. The compound rate of return can be approximated by $R_i \equiv \left(1 + r_{i,1}^p z\right) \cdot ... \cdot \left(1 + r_{i,H}^p z\right) - 1 = \sum_{h=1}^H r_{i,h}^p z + o\left(z\right)$. Capital income reads as $R_i a_{i,1} = \sum_{h=1}^H a_{i,h} r_{i,h}^p z + o\left(z\right)$. Therefore, when $z = 1/H$, the investment return r_i denotes the average return. Consider the setting in Section C, where the government taxes final wealth linearly according to τ_W.[20]

Information structure. As standard in the literature, assume that there are noise traders who have access to other investment technologies, such as human capital, or make random errors in their forecast of payoffs. The existence of noise traders prevents the full revelation of private information via the publicly-observed price and, as a result, a fully efficient financial market. Otherwise, nobody would have an incentive to purchase the private signal in the first place (Grossman-Stiglitz paradox). Accordingly, the net supply of risky assets, θ_h, is random. Assume that the net supply is normally distributed, $\theta_h \sim \mathcal{N}\left(\mathbb{E}\left(\theta\right), \sigma_\theta^2\right)$, and independent from payoffs. This technicality ensures that the equilibrium price is a noisy signal about the fundamentals of the economy.

Households can acquire financial knowledge, for example, by conducting research, obtaining financial education, or employing financial advisers. In particular, they observe a noisy private signal $s_{i,h} = b_h + \vartheta_{i,h}$ with $\vartheta_{i,h} \sim \mathcal{N}\left(0, \frac{1}{x_{i,h}}\right)$ and can purchase a signal precision of $x_{i,h} \in \mathbb{R}_+ \cup \{0\}$ at cost $v\left(x_{i,h}\right) z$, measured in monetary units, where $v\left(\cdot\right)$ is increasing, convex, twice continuously differentiable and $v\left(0\right) = 0$. That is, information acquisition becomes more and more costly. This assumption is in line with the idea that households obtain pieces of information, and each extra piece correlates with the previous ones. Nonetheless, this model gives rise to increasing returns to information acquisition. Moreover, assume that private signals are uncorrelated across households and that households cannot resell their information. As in reality, agency problems may constrain information resale or sharing.[21]

[20] Analyzing a nonlinear capital gains tax, $T_k\left(\cdot\right)$, with $T_k\left(0\right) = 0$, $T_k'\left(0\right) = 0$, and $T_k''\left(0\right) = 0$, leads to the same conclusions (see Appendix F.4). Therefore, the financial market, described here, also microfounds the formation of returns in the analysis of nonlinear taxes in Section G. Similarly, one can consider a linear capital gains tax as in Section 2.

[21] Observe the implicit assumption that knowledge fully depreciates intertemporally. Any departure from this

Timing. The timing of each subperiod is as follows. For a given amount of savings, households purchase financial knowledge $x_{i,h}$. Then, they observe the private and the public price signal. Households form rational expectations about the payoff of the risky asset given the observed signals and decide how much of their savings to invest in the risky asset. Formally, an investor i chooses a share of stocks, $\varsigma_{i,h}$, and a bond share, $(1-\varsigma_{i,h})$, given her expectation $\mathbb{E}_{i,h}\left(\cdot|\mathcal{F}_{i,h}\right)$ conditional on the information set $\mathcal{F}_{i,h}$ where $\mathcal{F}_{i,h}=\{s_{i,h},P_h\}$, if a signal has been acquired, and $\mathcal{F}_{i,h}=\{P_h\}$, else. Finally, payoffs realize.

Household problem. Given the portfolio choice $\varsigma_{i,h}$, the return of the portfolio reads as

$$r^p_{i,h}z=\varsigma_{i,h}\frac{\Pi_h-P_h}{P_h}+\left(1-\varsigma_{i,h}\right)r^f_h z \tag{38}$$

per unit of savings $a_{i,h-1}$. At the end of the subperiod household i's wealth is the portfolio's gross return net of costs of information acquisition

$$a_{i,h}=a_{i,h-1}\left(1+r^p_{i,h-1}z\right)-v\left(x_{i,h-1}\right)z.$$

I assume that the costs of information acquisition are monetary, realize at the end, and are deductible from the base of the capital tax.

Due to the model approximation used here, the main result that the portfolio return increases with wealth, derived in the next section, is robust to various permutations of these assumptions on the information costs. In particular, it does not matter when the monetary costs accrue. Moreover, when the costs of information acquisition are non-monetary, the key results will carry over with a minor constraint on the shape of the cost function.

Final wealth, $a_{i,H+1}$, can be recursively written as

$$a_{i,H+1}=a_{i,1}\left(1+\sum_{h=1}^{H}r^p_{i,h}z\right)-\sum_{h=1}^{H}v\left(x_{i,h}\right)z+o\left(z\right).$$

I assume that utility from final, after-tax wealth, $\underline{a}_{i,t+1}$, is linearly separable and isoelastic $u\left(\underline{a}_{i,t+1}\right)=\frac{\underline{a}_{i,t+1}^{1-\rho}-1}{1-\rho}$. Then, this utility is approximately given by $u\left[\left(1-\tau_W\right)\left(a_{i,1}\left(1+R_i\right)-v\left(X_i\right)\right)\right]+o\left(z\right)$ where $R_i\equiv\sum_{h=1}^{H}r^p_{i,h}z$ and $v\left(X_i\right)\equiv\sum_{h=1}^{H}v\left(x_{i,h}\right)z$. This justifies the preference structure in the dynamic economy of Section C.

It remains to show that $r_{i,h}=r\left(a_{i,h},\left\{a_{j,h}\right\}_{j\in[0,1]}\right)$. Firstly, note that utility from final wealth

assumption would, just as non-convex cost functions, strengthens the main results.

can also be written as

$$H \cdot u\left[(1-\tau_W)\left(a_{i,1}\left(1+r_{i,1}^{p}z\right)-v\left(x_{i,1}\right)z\right)\right]+o\left(1\right) \tag{39}$$

Hence, for a given distribution of initial wealth, $a_{i,1}$, the repeated financial market interaction in subperiod h is up to a constant *fully static*.[22] Therefore, in the following, I drop time indices in individual and aggregate variables for notational convenience. Accordingly, in each subperiod, households maximize their expected utility

$$\underset{x}{max}\ \mathbb{E}_i\left(\underset{\varsigma}{max}\ \mathbb{E}_i\left(u\left[(1-\tau_W)\left(a_i\left(1+r_i^{p}z\right)-v\left(x_i\right)z\right)\right]|\mathcal{F}_i\right)\right) \tag{40}$$

The set of optimal choices by household i on the financial market reads as $\{\varsigma_i, x_i\}$ which will be functions of initial savings. Moreover, denote the p.d.f. of savings as $g\left(a_i\right)$ and the c.d.f. as $G\left(a_i\right)$, respectively.

A side-effect of the model approximation is that one can rewrite the stochastic period utility in deterministic units

$$\begin{aligned}\mathbb{E}_i\left(u\left[(1-\tau_W)\left(a_i\left(1+r_i^{p}z\right)-v\left(x_i\right)z\right)\right]\right) &= u\left[(1-\tau_W)\left(a_i\left(1+\mathbb{E}\left(r_i^{p}z\right)\right)-v\left(x_i\right)z\right)\right]\\ &+\frac{1}{2}u''\left[(1-\tau_W)\,a_i\right](1-\tau_W)^2\,\mathbb{V}\left(a_i r_i^{p}z\right)+o\left(z\right).\end{aligned}$$

To get this expression, approximate the expected utility around the mean portfolio return. Thus, second-period utility features a deterministic mean-variance trade-off in the spirit of Markowitz (1952, 1959). Households trade off endogenous ex ante risk and returns. I derive these measures in the following. Therefore, the tax analysis with deterministic returns is sufficient.

Aggregate variables. Denote the risk tolerance of a household, who invests a_i, as the inverse coefficient of absolute risk aversion $\psi\left(a_i\right) \equiv \frac{-u'(a_i)}{u''(a_i)}$. With the specified utility function, $\psi\left(a_i\right) = a_i/\rho$. In principle, $\psi'\left(a_i\right) > 0$ would be sufficient to obtain scale dependence.

Moreover, dropping the time index on the aggregate variables, define the aggregate risk-taking by $\mathcal{T} \equiv \int_i \mathcal{T}_i di \equiv \int_i \psi\left(a_i\right)di$, the aggregate noisiness by $\mathcal{N} \equiv \int_i \mathcal{N}_i di \equiv \int_i \frac{\psi(a_i)}{h_0(\mathcal{I})+x_i}di$, and the aggregate informativeness of the price by $\mathcal{I} \equiv \int_i \mathcal{I}_i di \equiv \frac{1}{\sigma^2}\int_i \frac{x_i\psi(a_i)}{h_0(\mathcal{I})+x_i}di$, where $h_0\left(\mathcal{I}\right) \equiv \frac{1}{\sigma_b^2}+\frac{\mathcal{I}^2}{\sigma_\theta^2}$ measures the precision of the public signal. Therefore, the variable $\mathcal{T}$ aggregates risk tolerance or risk-taking of all households. $\mathcal{N}$ summarizes the noisiness (inverse precision) of the prior, the

[22] Incorporating dynamic aspects, e.g., the accumulation of wealth and the resulting spread in the wealth distribution, would only strengthen the main result that wealthier households obtain higher rates of return than poorer ones.

stock price, and the private signals of households, whereas $\mathcal{I}$ measures the total signal precision relative to the total precision. Both $\mathcal{N}$ and $\mathcal{I}$ are weighted by the risk tolerance. The definition of these three variables will prove convenient when deriving the equilibrium of the economy. Now, one can define the rational expectations equilibrium of the financial market.

Rational expectations equilibrium. Define a rational expectations equilibrium as the set of choices $\{\varsigma_i, x_i\}$, the stock's price as a function of Π and θ and the informativeness $\mathcal{I}$ such that

(1) households optimally choose their portfolio and signal precision

$$\varsigma_i = \varsigma\left(S_i, x_i, a_i; P, \mathcal{I}\right) \equiv arg\max_{\varsigma} \mathbb{E}_i\left(u\left[\left(1-\tau_W\right)\left(a_i\left(1+r_i^p z\right) - v\left(x_i\right) z\right)\right] | \mathcal{F}_i\right) \tag{41}$$

and

$$x_i = x\left(a_i; \mathcal{I}\right) \equiv arg\max_{x} \mathbb{E}_i\left[\max_{\varsigma} \mathbb{E}_i\left(u\left[\left(1-\tau_W\right)\left(a_i\left(1+r_i^p z\right) - v\left(x_i\right) z\right)\right] | \mathcal{F}_i\right)\right], \tag{42}$$

(2) P clears the stock market

$$\int_i \frac{\varsigma_i a_i}{P} di = \theta, \tag{43}$$

and

(3) the implied informativeness of the price is consistent with observed choices of individual information precision

$$\mathcal{I} = \frac{1}{\sigma^2}\int_i \frac{x\left(a_i; \mathcal{I}\right)\psi\left(a_i\right)}{h_0\left(\mathcal{I}\right) + x\left(a_i; \mathcal{I}\right)} di. \tag{44}$$

E.2 The Equilibrium

In the following, I show that, in the approximated Grossman and Stiglitz (1980) economy, investment returns and their distribution depend on capital justifying the reduced form assumption on the capital gains functional in the sections before. I solve the model by backward induction. First, one shows that there exists a log-linear rational expectations equilibrium and derive portfolio choices and the equilibrium stock price. Then, to demonstrate that the amount of information acquisition, x_i, increases in the portfolio size, a_i, one characterizes the demand for information by the first-order condition

$$v'\left(x_i\right) = \frac{1}{2\rho} a_i \mathcal{S}'\left(x_i; \mathcal{I}\right), \tag{45}$$

where $\mathcal{S}\left(x_i; \mathcal{I}\right)$ is the expected squared Sharpe ratio of an investor. Wealthy investors purchase more information than poorer ones. There exists a threshold value $a_i^*\left(\mathcal{I}\right)$ below which nobody obtains information. There is a congestion effect. The threshold wealth $a_i^*\left(\mathcal{I}\right)$ is increasing in $\mathcal{I}$. Hence, a rise in the aggregate informativeness lowers the number of investors who choose to purchase information.

Furthermore, note that information is a strategic substitute. That is, $x(a_i; \mathcal{I})$ is a decreasing function of $\mathcal{I}$. The higher the informativeness of the public signal (price), the lower is the need for acquiring private information. In other words, the information acquisition by all investors imposes an externality on an individual investor via the equilibrium price. Investors do not internalize this externality. Finally, it can be shown that there exists a unique scalar for $\mathcal{I}$ and, thus, for $\mathcal{N}$. Therefore, the log-linear equilibrium is unique.

E.2.1 Portfolio Returns and Sharpe Ratio

Now, I present the implications of information acquisition for portfolio returns. As we have seen, wealthier investors acquire more information, even though each extra piece of information becomes more and more costly. Does this information advantage help investors to generate higher rates of return? To answer this question, define the excess return of investor i's portfolio $r_i^{pe}z \equiv r_i^p z - r^f z$.

Lemma 1 (Returns, variance, and Sharpe ratio)**.** *The expected excess return, its variance, and the Sharpe ratio are increasing in x_i which rises in a_i:*

$$\mathbb{E}\left(r_i^{pe}z\right) = \mathbb{E}\left(r_i^p z\right) - r^f z = \frac{1}{\rho}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o\left(z\right), \tag{46}$$

$$\mathbb{V}\left(r_i^{pe}z\right) = \mathbb{V}\left(r_i^p z\right) = \frac{1}{\rho^2}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o\left(z\right), \tag{47}$$

and

$$\frac{\mathbb{E}\left(r_i^{pe}z\right)}{\sqrt{\mathbb{V}\left(r_i^{pe}z\right)}} = \sqrt{\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z} + o\left(1\right). \tag{48}$$

Proof. See Appendix F.2. □

Lemma 1 reveals how the portfolio returns (and its risk) relate to the individual's signal precision x_i, portfolio sizes a_i, and the relative risk aversion $1/\rho$. Both the expected excess return and its standard deviation are declining in the relative risk aversion. Moreover, these variables increase in the degree of individual information that rises in the portfolio size. Hence, wealthier investors obtain higher returns and are willing to take more risk relative to poorer households. Moreover, returns depend on aggregate information.

To sum up, an individual's demand for stocks and information, as well as her (risk-adjusted) return, depend on her amount of investment and, through the equilibrium price, on others' investments. Households become richer because they are rich. As a result, the final wealth distribution is more unequal than the initial one. This insight originates from Arrow (1987).

Moreover, an investor's return does not directly depend on her capital tax. This feature derives from the linear approximation of the economy and the CRRA utility function. Altogether, this financial market interaction justifies the reduced form assumption on the endogenous return inequality in Section C and G, $r_i\left(a_i,\{a_j\}_{j\in[0,1]}\right)$.

E.2.2 An Example

Suppose, for simplicity, that $\mathbb{E}(\theta)=0$ and $v(x_i)=\kappa x_i$. Due to the linearity of costs, the rents from private signal extraction are constant conditional on a given amount of investment. A higher degree of public information reduces one-to-one the demand for private information. Moreover, let $a_0>a_i^*(\mathcal{I})$. Then, the elasticity of the return (in a given subperiod) with respect to the amount of investment is positive

$$\varepsilon_i^{\mathbb{E}(r^p z),a}\equiv\frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log\left(a_i\right)}=\frac{\sqrt{\rho\kappa/\left(2\sigma^2 a_i\right)}}{\mathcal{S}\left(a_i,\{a_j\}_{j\in[0,1]}\right)/\rho+r^f}>0.$$

Also, note that the expected return is concave in the amount of investment. Therefore, own-return elasticity decreases with a_i.

The cross-return elasticity reads as

$$\gamma_{i,i'}^{\mathbb{E}(r^p z),a}\equiv\frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log\left(a_{i'}\right)}=\sum_{\mathcal{A}\in\{\mathcal{T},\mathcal{N},\mathcal{I}\}}\frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial\mathcal{A}}\frac{\partial\mathcal{A}_{i'}}{\partial log\left(a_{i'}\right)}.$$

One can show that by the linearity of the cost function it is multiplicatively separable. That is, $\gamma_{i,i'}^{\mathbb{E}(r^p z),a}=\frac{1}{\mathbb{E}\left(r_i^p z\right)}\delta_{i'}^{\mathbb{E}(r^p z),a}$.

Observe that the cross-return elasticity carries risk and information externalities. Investors are rewarded for risk that they are willing to take on the stock market. The variability of the price measures this risk: $\mathbb{V}\left(log\left(P\right)\right)=\mathbb{V}\left(p_\xi\xi\right)$ where ξ is the public signal and p_ξ is the responsiveness of the price to the public signal. Two channels affect the amount of this aggregate risk and, as a result, individual returns.

Firstly, a rise in aggregate information, $\mathcal{I}$, lowers the variance of ξ and, therefore, lowers portfolio returns. Secondly, the sensitivity of the stock price to the price signal, p_ξ, is determined in general equilibrium. As the aggregate noisiness, $\mathcal{N}$, declines, the equilibrium stock price becomes more sensitive to the price signal so that p_ξ increases. Similarly, a rise in risk tolerance, $\mathcal{T}$, increases the demand for stocks intensifying the relation between the price and the public signal. Hence, a rise in $\mathcal{T}$ (a reduction in $\mathcal{N}$) increases the variability of the public signal.

Altogether, a rise in portfolio size $a_{i'}$ (and, therefore, in information $x_{i'}$) has opposing effects on the return of household i. For simplicity, let $\sigma^2 = \sigma_b^2 = \sigma_\theta^2 = 1$. Then, one can show that $\delta_{i'}^{\mathbb{E}(r^p z),a} \geq 0$ for $a_{i'} \leq \tilde{a}$ and $\delta_{i'}^{\mathbb{E}(r^p z),a} < 0$ for $a_{i'} > \tilde{a}$. Whereas an investor's marginal contribution to risk is constant, contributions to information are nonlinear in the amount of investment. For instance, the impact of wealthy investors on information is larger than the one of poorer investors (i.e., $\frac{\partial^2 \mathcal{I}_{i'}}{\partial a_{i'}^2} > 0$). They contribute marginally more to the level of aggregate information, which reduces uncertainty and, hence, the idiosyncratic reward for risk ($\mathbb{E}\left(r_i^p z\right)$).

Consequently, this setting features trickle-up. Consider a tax cut on the wealth of the rich. As a reaction, wealthy investors increase their portfolio size which allows them to generate higher rates of return because they acquire more information ($\varepsilon_i^{\mathbb{E}(r^p z),a} > 0$). At the same time, the level of aggregate information increases. As a consequence, the value of private information decreases. The reward for the small amount of private information, that poorer households acquire, declines ($\delta_{i'}^{\mathbb{E}(r^p z),a} < 0$). Therefore, the tax cut shifts capital income from the bottom to the top.

Of course, this observation holds when all households, even the poor, invest in financial knowledge (i.e., $a_0 > a_i^*\left(\mathcal{I}\right)$). Suppose that $a_i = a_i^*\left(\mathcal{I}\right)$ for some $i \in (0,1)$. Then, the poor, who do not invest in information, may benefit from a tax cut for the rich, as they only rely on public information. In this situation, only the middle class suffers from a loss in their rents from private information acquisition.

E.3 Extensions

In this section, I extend the financial market model by considering two practically relevant modifications of the financial market model. First, I consider career effects. In the second extension, I deal with type dependence. Throughout this section, suppose the assumptions from the example hold. That is, let $\mathbb{E}\left(\theta\right) = 0$, $\sigma^2 = \sigma_b^2 = \sigma_\theta^2 = 1$, and $v\left(x_i\right) = \kappa x_i$. Moreover, assume that $a_0 > a_i^*\left(\mathcal{I}\right)$.

E.3.1 Career Effects

Wealthy households may not only obtain high financial knowledge since their portfolios are sizable but also because of the professional network they build during their career. In other words, as they earn a high income and, as a result, become wealthy, they gain access to specialist knowledge about financial markets either because they work in the finance industry or they get to know financial experts. This channel additionally boosts their portfolio returns.

To formalize this, let $v\left(x_i, y_i\right)$ where $\frac{\partial^2 v(x_i,y_i)}{\partial y_i^2} > 0$ and $\frac{\partial v(x_i,y_i)}{\partial x_i \partial y_i} < 0$. The marginal costs of purchasing information decrease with an individual's income y_i. Then, the Sharpe ratio

$$\mathcal{S}\left(a_i, l_i, \{a_j\}_{j\in[0,1]}, \{l_j\}_{j\in[0,1]}\right)$$

and, accordingly, the expected rate of return, as well as its variance, increase with an individual's labor supply.[23] As labor supply increases with i, this force amplifies the main feature of the model of endogenous return inequality. Put differently, $\varepsilon_i^{\mathbb{E}(r^p z),l} \equiv \frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log(l_i)} > 0$. In general equilibrium, $\gamma_{i,i'}^{\mathbb{E}(r^p z),l} \equiv \frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log(l_{i'})} \neq 0$.

E.3.2 Type Dependence

As noted in the literature on inequality (e.g., Benhabib, Bisin, and Zhu (2011)), type dependence explains the thick tail in the distribution of wealth observed in many countries. Applied to the financial market setting, this refers to a situation where the rich are also talented in investing their money.

The easiest way to incorporate type dependence is to let κ_i vary by type. That is, suppose κ_i is decreasing in the index i. Thus, there is heterogeneity not only in hourly wages, but also in the marginal costs of information acquisition. Accordingly, an investor's Sharpe ratio $\mathcal{S}_i(\cdot)$ is indexed by i. The presence of cost heterogeneity amplifies the inequality in returns. The reasoning is as follows. Wealthy, talented investors acquire more financial knowledge than without type dependence, as it is cheaper for them. Therefore, they earn higher returns. In turn, the incentives to save rise such that their portfolio increases in size. Because of scale dependence, this further boosts their returns.

Moreover, the distribution of own-return elasticities is affected. To see this, compare own-return semi-elasticities of household i and j where $i > j$: $\frac{\mathbb{E}\left(r_i^p z\right)\varepsilon_i^{\mathbb{E}(r^p z),a}}{\mathbb{E}\left(r_j^p z\right)\varepsilon_j^{\mathbb{E}(r^p z),a}} = \sqrt{\frac{a_j}{a_i}}\sqrt{\frac{\kappa_i}{\kappa_j}}$. There are two effects that compress the distribution of own-return semi-elasticities. Firstly, $\sqrt{\frac{\kappa_i}{\kappa_j}} < 1$. Secondly, type dependence leads to more return inequality which boosts wealth inequality. Thus, $\sqrt{\frac{a_j}{a_i}}$ is lower in the presence of type dependence. The effect on the distribution of own-return elasticities is even larger $\frac{\varepsilon_i^{\mathbb{E}(r^p z),a}}{\varepsilon_j^{\mathbb{E}(r^p z),a}} = \frac{\mathbb{E}\left(r_j^p z\right)}{\mathbb{E}\left(r_i^p z\right)}\sqrt{\frac{a_j}{a_i}}\sqrt{\frac{\kappa_i}{\kappa_j}}$ because return inequality, $\frac{\mathbb{E}\left(r_j^p z\right)}{\mathbb{E}\left(r_i^p z\right)}$, directly enters the expression. Therefore, the presence of type dependence compresses the distribution of own-return elasticities.

In general equilibrium, the distribution of cross-return semi-elasticities is unaffected by type

[23] $\frac{\partial v(x_i,y_i)}{\partial x_i \partial y_i} < 0$ implies by the second fundamental theorem of calculus that $\frac{\partial v(x_i,y_i)}{\partial y_i} \neq 0$. Therefore, the labor supply elasticities are modified by an additional marginal effect on information costs.

dependence, whereas the effect on the distribution of cross-return elasticities depends on the effects on return inequality $\frac{\gamma_{i,i'}^{\mathbb{E}(r^p z),a}}{\gamma_{j,i'}^{\mathbb{E}(r^p z),a}} = \frac{\mathbb{E}\left(r_j^p z\right)}{\mathbb{E}\left(r_i^p z\right)}$. If type dependence triggers a rise in return inequality, the distribution of cross-return elasticities will flatten.

F Proofs of Section E

F.1 Approximations

I show, by induction, that the statement $P(H) : \Pi_{h=1}^{H}\left(1 + r_{i,h}^p z\right) = 1 + \sum_{h=1}^{H} r_{i,h}^p z + o(z)$ holds for any $H \geq 1$. The base case, $P(1)$, is trivially fulfilled. For the inductive step, let $P(k)$ hold. Then, $P(k+1)$ is also true since

$$\Pi_{h=1}^{k}\left(1 + r_{i,h}^p z\right) \cdot \left(1 + r_{i,k+1}^p z\right) = \left(1 + \sum_{h=1}^{k} r_{i,h}^p z + o(z)\right)\left(1 + r_{i,k+1}^p z\right)$$
$$= 1 + \sum_{h=1}^{k} r_{i,h}^p z + \left(1 + \sum_{h=1}^{k} r_{i,h}^p z\right) r_{i,k+1}^p z + o(z) = 1 + \sum_{h=1}^{k+1} r_{i,h}^p z + o(z).$$

Using this expression, period-h wealth can be written as

$$a_{i,h} = a_{i,h-1}\left(1 + r_{i,h-1}^p z\right) - v\left(x_{i,h-1}\right) z = \left[a_{i,h-2}\left(1 + r_{i,h-2}^p z\right) - v\left(x_{i,h-2}\right) z\right]\left(1 + r_{i,h-1}^p z\right) - v\left(x_{i,h-1}\right) z$$
$$= a_{i,h-2}\left(1 + r_{i,h-1}^p z + r_{i,h-2}^p z\right) - v\left(x_{i,h-2}\right) z - v\left(x_{i,h-1}\right) z + o(z) = \ldots$$
$$= a_{i,1}\left(1 + \sum_{j=1}^{h-1} r_{i,h-j}^p z\right) - \sum_{j=1}^{h-1} v\left(x_{i,h-j}\right) z + o(z) = a_i\left(1 + \sum_{s=1}^{h-1} r_{i,s}^p z\right) - \sum_{s=1}^{h-1} v\left(x_{i,s}\right) z + o(z)$$

for any $h = 1, \ldots, H+1$. Capital income is given by

$$R_i a_{i,1} = \sum_{h=1}^{H} a_{i,1} r_{i,h}^p z + o(z) = \sum_{h=1}^{H} a_{i,h} r_{i,h}^p z + \sum_{h=1}^{H}\left(a_{i,1} - a_{i,h}\right) r_{i,h}^p z + o(z) = \sum_{h=1}^{H} a_{i,h} r_{i,h}^p z + o(z).$$

Defining the overall information effort as $x_i \equiv \sum_{h=1}^{H} x_{i,h} z$, the information costs can be approximated by

$$v\left(X_i\right) \equiv v\left(\sum_{h=1}^{H} x_{i,h} z\right) \equiv v\left(x_{i,1} z, \ldots, x_{i,H} z\right) = v(0, \ldots, 0) + \sum_{h=1}^{H} \frac{\partial v(0, \ldots, 0)}{\partial x_{i,h}}\left(x_{i,h} z - 0\right) + o(z)$$
$$= v'(0)\, x_{i,1} z + \ldots + v'(0)\, x_{i,H} z + o(z) = \sum_{h=1}^{H}\left(v\left(x_{i,h}\right) z - v(0)\right) + o(z) = \sum_{h=1}^{H} v\left(x_{i,h}\right) z + o(z)$$

Therefore, one can rewrite the utility from final wealth as

$$u\left(a_{i,H+1}\right)=u\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+\sum_{h=1}^{H}r_{i,h}^{p}z\right)-\sum_{h=1}^{H}v\left(x_{i,h}\right)z\right)\right]+o\left(z\right)$$
$$=u\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+R_{i}\right)-v\left(X_{i}\right)\right)\right]+o\left(z\right),$$

which justifies the preference structure in Section C. Alternatively, one can express the utility from final wealth as

$$u\left(a_{i,H+1}\right)=u\left(a_{i,1}+\sum_{h=1}^{H}\Delta a_{h}\right)+o\left(z\right)=u\left(\Delta a_{1},...,\Delta a_{H}\right)+o\left(z\right)=u\left(0,...,0\right)+\sum_{h=1}^{H}\frac{\partial u\left(0,...,0\right)}{\partial\Delta a_{h}}\Delta a_{h}+o\left(z\right)$$
$$=u\left(a_{i,1}\right)+\sum_{h=1}^{H}\left(u\left(a_{i,1}+\Delta a_{h}\right)-u\left(a_{i,1}\right)\right)+o\left(z\right)=\sum_{h=1}^{H}u\left(a_{i,1}\left(1+r_{i,h}^{p}z\right)-v\left(x_{i,h}\right)z\right)+o\left(1\right),$$

where I defined $\Delta a_{h}\equiv a_{i,1}r_{i,h}^{p}z-v\left(x_{i,h}\right)z$. By the simplifying assumption that knowledge fully depreciates intertemporally, Equation (39) follows.

As I will show later, any moment higher than the return variance is negligible. Accordingly, expected period-utility can be approximated around the utiltity from expected wealth as follows

$$\mathbb{E}_{i}\left(u\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+r_{i,1}^{p}z\right)-v\left(x_{i,1}\right)z\right)\right]\right)=\mathbb{E}_{i}\left(u\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+\mathbb{E}\left(r_{i,1}^{p}z\right)\right)-v\left(x_{i,1}\right)z\right)\right]\right)$$
$$+\left(1-\tau_{W}\right)\mathbb{E}_{i}\left(u'\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+\mathbb{E}\left(r_{i,1}^{p}z\right)\right)-v\left(x_{i,1}\right)z\right)\right]\left[a_{i,1}r_{i,1}^{p}z-a_{i,1}\mathbb{E}\left(r_{i,1}^{p}z\right)\right]\right)$$
$$+\frac{1}{2}\left(1-\tau_{W}\right)^{2}\mathbb{E}_{i}\left(u''\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+\mathbb{E}\left(r_{i,1}^{p}z\right)\right)-v\left(x_{i,1}\right)z\right)\right]\left[a_{i,1}r_{i,1}^{p}z-a_{i,1}\mathbb{E}\left(r_{i,1}^{p}z\right)\right]^{2}\right)+o\left(z\right)$$
$$=u\left[\left(1-\tau_{W}\right)\left(a_{i,1}\left(1+\mathbb{E}\left(r_{i,1}^{p}z\right)\right)-v\left(x_{i,1}\right)z\right)\right]+\frac{1}{2}u''\left[\left(1-\tau_{W}\right)a_{i,1}\right]\left(1-\tau_{W}\right)^{2}\mathbb{V}\left(a_{i,1}r_{i,1}^{p}z\right)+o\left(z\right).$$

F.2 The Financial Market Equilibrium and Linear Taxation

Equilibrium price, existence, and demand for stocks. In the following, I characterize the financial market equilibrium in subperiod 1 (and, therefore, for each subperiod h). Therefore, I completely drop time indices in this section. I start with portfolio choices and derive the equilibrium stock price. Lemma 2 summarizes the results.

Lemma 2 (Existence of equilibrium, equilibrium price, and portfolio choice). *Assume z is small. Then, there exists a log-linear rational expectations equilibrium. The equilibrium price is linear in $\xi\equiv b-\frac{1}{\mathcal{I}}\theta$*

$$log\left(P\right)=pz=\left(p_{0}+p_{\xi}\xi-r^{f}\right)z+o\left(z\right) \tag{49}$$

where $p_{0}\equiv\frac{\mathcal{N}}{\mathcal{T}}\left[\frac{\mathbb{E}(b)}{\sigma_{b}^{2}}+\frac{\mathcal{I}\mathbb{E}(\theta)}{\sigma_{\theta}^{2}}\right]+\frac{1}{2}\sigma^{2}$ and $p_{\xi}\equiv1-\frac{\mathcal{N}}{\mathcal{T}\sigma_{b}^{2}}$. The optimal investment in the risky asset is *given by*

$$\varsigma_i = \frac{1}{\rho\sigma\sqrt{z}}\lambda_i + o\left(1\right) \tag{50}$$

where $\lambda_i \equiv \frac{\sqrt{z}}{\sigma}\left[\frac{1}{h_0(\mathcal{I})+x_i}\left(\frac{\mathbb{E}(b)}{\sigma_b^2}+\frac{\mathcal{I}\mathbb{E}(\theta)}{\sigma_\theta^2}+\frac{\mathcal{I}^2}{\sigma_\theta^2}\xi+x_i s_i\right)+\frac{1}{2}\sigma^2-p-r^f\right]$ *is the investor's Sharpe ratio.*

The proof of Lemma 2 involves three steps. Conjecturing the log-linear equilibrium price (Equation (49)), determine the conditional variance and expectation of payoffs (step 1), derive the optimal portfolio (step 2), and determine the equilibrium price using the stock market clearing confirming the price conjecture (step 3).

Step 1: By the law of total conditional variance and expectation, the conditional variance of payoff and the conditional expected payoff read as

$$\begin{aligned}\mathbb{V}_i\left(\pi z|\mathcal{F}_i\right) &= \mathbb{E}_i\left(\mathbb{V}_i\left(\pi z|b,\mathcal{F}_i\right)|\mathcal{F}_i\right)+\mathbb{V}_i\left(\mathbb{E}_i\left(\pi z|b,\mathcal{F}_i\right)|\mathcal{F}_i\right)\\ &= \mathbb{E}_i\left(\mathbb{V}_i\left(\pi z|b\right)|\mathcal{F}_i\right)+\mathbb{V}_i\left(bz|\mathcal{F}_i\right)=\sigma^2 z+o\left(z\right)\end{aligned}$$

and, using $b \equiv \xi + \frac{1}{\mathcal{I}}\theta$ in Lemma 2,

$$\begin{aligned}\mathbb{E}_i\left(\pi z|\mathcal{F}_i\right) &= \mathbb{E}_i\left(\mathbb{E}_i\left(\pi z|b,\mathcal{F}_i\right)|\mathcal{F}_i\right)=\mathbb{E}_i\left(bz|\mathcal{F}_i\right)\\ &= \frac{1}{h_0\left(\mathcal{I}\right)+x_i}\left[\frac{1}{\sigma_b^2}\mathbb{E}\left(b\right)+\frac{\mathcal{I}}{\sigma_\theta^2}\mathbb{E}\left(\theta\right)+\frac{\mathcal{I}^2}{\sigma_\theta^2}\xi+x_i s_i\right]z+o\left(z\right).\end{aligned}$$

Step 2: In the following, I approximate the household's Euler equation

$$\begin{aligned}0 &= \mathbb{E}_i\left[u'\left(\left(1-\tau_W\right)\left(a_i\left(1+\varsigma_i\frac{\Pi-P}{P}+\left(1-\varsigma_i\right)r^f z\right)-v\left(x_i\right)z\right)\right)\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right]\\ &= u'\left(\left(1-\tau_W\right)a_i\right)\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right]\\ &+\left(1-\tau_W\right)a_i\varsigma_i u''\left(\left(1-\tau_W\right)a_i\right)\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)^2|\mathcal{F}_i\right]+o\left(z\right)\end{aligned} \tag{51}$$

that determines the optimal portfolio choice. Note that

$$\begin{aligned}\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right] &= \mathbb{E}_i\left[\left(\frac{exp\left(\pi z\right)-exp\left(pz\right)}{exp\left(pz\right)}-r^f z\right)|\mathcal{F}_i\right]\\ &= \mathbb{E}_i\left[\left(\frac{1+\pi z+\frac{1}{2}\left(\pi z\right)^2+o\left(z^2\right)-1-pz-\frac{1}{2}\left(pz\right)^2-o\left(z^2\right)}{1+pz+o\left(z\right)}\right)|\mathcal{F}_i\right]-r^f z\\ &= \mathbb{E}_i\left(\pi z|\mathcal{F}_i\right)+\frac{1}{2}\mathbb{E}_i\left[\left(\left(\pi z\right)^2-\left(pz\right)^2\right)|\mathcal{F}_i\right]-pz-r^f z+o\left(z\right)\\ &= \mathbb{E}_i\left(bz|\mathcal{F}_i\right)+\frac{1}{2}\sigma^2 z-pz-r^f z+o\left(z\right)\end{aligned} \tag{52}$$

and

$$\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)^2|\mathcal{F}_i\right]=\mathbb{E}_i\left[\left(\pi z+\frac{1}{2}(\pi z)^2-pz-\frac{1}{2}(pz)^2-r^f z\right)^2|\mathcal{F}_i\right]+o(z)$$
$$=\mathbb{E}_i\left[(\pi z)^2|\mathcal{F}_i\right]+o(z)=\sigma^2 z+o(z). \tag{53}$$

Plug these expressions and the conjectured equilibrium price into Equation (51). To get Equation (50), rearrange the resulting expression and observe that $\frac{-u'((1-\tau_W)a_i)}{(1-\tau_W)a_i u''((1-\tau_W)a_i)}=\frac{1}{\rho}$.

Step 3: Plug Equation (50) and the definitions of the aggregate variables into the stock market clearing condition (Equation (41)) to get

$$\theta=\frac{1}{\sigma^2}\left[\left(\frac{1}{\sigma_b^2}E(b)+\frac{\mathcal{I}}{\sigma_\theta^2}E(\theta)+\frac{\mathcal{I}^2}{\sigma_\theta^2}\left(b-\frac{1}{\mathcal{I}}\theta\right)\right)\mathcal{N}+b\sigma^2\mathcal{I}+\mathcal{T}\left(\frac{1}{2}\sigma^2-p-r^f\right)\right]+o(1).$$

Rearrange to conclude that Equation (49) is fulfilled.

Demand for information and equilibrium uniqueness. In Lemma 3, I characterize the demand for information and confirm the uniqueness of the equilibrium.

Lemma 3 (Demand for information, equilibrium informativeness, and uniqueness of equilibrium)**.** *Assume z is small. There exists a threshold wealth $a_i^*(\mathcal{I})\equiv 2\rho v'(0)\sigma^2 h_0(\mathcal{I})^2$, above which there is positive information acquisition, x_i, that increases in a_i according to the first-order condition*

$$v'(x_i)=\frac{a_i}{2\rho}\mathcal{S}'(x_i;\mathcal{I}), \tag{54}$$

where $\mathcal{S}(x_i;\mathcal{I},\mathcal{N},\mathcal{T})z=\mathbb{E}_i(\lambda_i^2)$ is expected squared Sharpe ratio of an investor. $\mathcal{S}(x_i;\mathcal{I},\mathcal{N},\mathcal{T})$ is increasing and concave in the precision of the private signal x_i. Therefore, the informativeness of the price can be written as

$$\mathcal{I}=\frac{1}{\rho\sigma^2}\int_{a_i^*(\mathcal{I})}^{a_1}\frac{a_i x_i(a_i;\mathcal{I})}{h_0(\mathcal{I})+x_i(a_i;\mathcal{I})}dG(a_i). \tag{55}$$

There exists a unique log-linear equilibrium.

To proof Lemma 3, observe that a households expected squared Sharpe ratio is given by

$$\mathcal{S}(x_{i,1};\mathcal{I})z\equiv\mathbb{E}_i\left(\lambda_i^2\right)=\mathbb{V}_i(\lambda_i)+\mathbb{E}_i(\lambda_i)^2$$
$$=-\frac{z}{\sigma^2}\frac{1}{h_0(\mathcal{I})+x_i}+\frac{z}{\sigma^2}\left[\frac{\sigma_\theta^2}{\mathcal{I}^2}p_\xi^2+\sigma_b^2(1-p_\xi)^2+\frac{E(\theta)^2}{\mathcal{I}^2}\left(1-h_0(\mathcal{I})\frac{\mathcal{N}}{\mathcal{T}}\right)^2\right]. \tag{56}$$

Similar to above, one approximates

$$\begin{aligned}\mathbb{E}_i\left[u\left(\cdot\right)|\mathcal{F}_i\right] = {} & u\left(\left(1-\tau_W\right)a_i\right)+\left(1-\tau_W\right)u'\left(\left(1-\tau_W\right)a_i\right)\\ & \cdot\left[a_i\varsigma_i\left(\mathbb{E}_i\left(\pi z|\mathcal{F}_i\right)+\frac{1}{2}\mathbb{V}_i\left(\pi z|\mathcal{F}_i\right)-pz-r^f z\right)+a_i r^f z-v\left(x_i\right)z\right]\\ & +\frac{1}{2}\left(1-\tau_W\right)^2 u''\left(\left(1-\tau_W\right)a_i\right)a_i^2\varsigma_i^2\mathbb{V}_i\left(\pi z|\mathcal{F}_i\right)+o\left(z\right)\end{aligned}$$

to obtain a non-stochastic expression for

$$\mathbb{E}_i\left[u\left(\cdot\right)\right]=u\left(\left(1-\tau_W\right)a_i\right)+\left(1-\tau_W\right)u'\left(\left(1-\tau_W\right)a_i\right)z\left(\frac{a_i}{2\rho}\frac{\mathbb{E}_i\left(\lambda_i^2\right)}{z}+a_i r^f-v\left(x_i\right)\right)+o\left(z\right)$$

Then, optimize over signal precision x_i. The first-order condition

$$v'\left(x_i\right)=\frac{1}{2\rho}a_i\mathcal{S}'\left(x_i;\mathcal{I}\right)=\frac{a_i}{2\rho\sigma^2}\left(\frac{1}{\sigma_b^2}+\frac{\mathcal{I}^2}{\sigma_\theta^2}+x_i\right)^{-2}$$

is sufficient by the second-order condition

$$\frac{\partial^2\mathbb{E}_i\left(u\left(\cdot\right)\right)}{\partial x_i^2}=\frac{1}{2\rho}a_i\mathcal{S}''\left(x_i;\mathcal{I}\right)-v''\left(x_i\right)<0$$

and, hence, characterizes the unique solution to the household information acquisition problem.

By the implicit function theorem, information acquisition rises with wealth

$$\frac{dx_i}{da_i}\propto\frac{\partial^2\mathbb{E}_i\left(u\left(\cdot\right)\right)}{\partial x_i\partial a_i}=\frac{1}{2\rho}\mathcal{S}'\left(x_i;\mathcal{I}\right)>0$$

and $a_i^*\left(\mathcal{I}\right)\equiv 2\rho v'\left(0\right)\sigma^2\mathcal{S}\left(0;\mathcal{I}\right)^{-1}=2\rho v'\left(0\right)\sigma^2 h_0\left(\mathcal{I}\right)^2$ is the threshold wealth level above which there is information acquisition. Denote i^* as the respective threshold household. Again, use the implicit function theorem to show that

$$\frac{dx_i}{d\mathcal{I}}\propto\frac{\partial^2\mathbb{E}_i\left(u\left(\cdot\right)\right)}{\partial x_i\partial\mathcal{I}}=\frac{\partial\mathcal{S}'\left(x_i;\mathcal{I}\right)}{\partial\mathcal{I}}=-\frac{2a_i}{\rho\sigma^2\left(\frac{1}{\sigma_b^2}+\frac{\mathcal{I}^2}{\sigma_\theta^2}+x_i\right)^3}\frac{\mathcal{I}}{\sigma_\theta^2}<0.$$

Finally, one needs to show that the equilibrium information $\mathcal{I}$ is uniquely determined for a given distribution of wealth. Define

$$\sum\left(\mathcal{I}\right)\equiv\mathcal{I}-\frac{1}{\rho\sigma^2}\int_{i^*}^{1}\frac{a_i x_i\left(a_i;\mathcal{I}\right)}{h_0\left(\mathcal{I}\right)+x_i\left(a_i;\mathcal{I}\right)}di.$$

One can demonstrate that the differential of this expression is positive

$$\frac{d\sum(\mathcal{I})}{d\mathcal{I}} = 1 - \frac{1}{\rho\sigma^2}\int_{i^*}^{1} \frac{h_0(\mathcal{I})\frac{dx_i(a_i;\mathcal{I})}{d\mathcal{I}} - x_i(a_i;\mathcal{I})\frac{dh_0(\mathcal{I})}{d\mathcal{I}}}{\left(h_0(\mathcal{I}) + x_i(a_i;\mathcal{I})\right)^2} di > 0.$$

Moreover, $\sum(0) \leq 0$, since $x_i(a_i;0) \geq 0$, and $\sum(\infty) \geq 0$, as $x_i(a_i;\infty) = 0$. By the continuity of $\sum(\mathcal{I})$, there is a unique $\mathcal{I}$ such that $\sum(\mathcal{I}) = 0$. Therefore, $\mathcal{N}$ and $\mathcal{T}$ are also uniquely defined.

Returns, variance, and Sharpe ratio. Lastly, I derive the key moments of return rates conditional on the amount of investment. Excess portfolio returns are given by $r_i^{pe}z \equiv r_i^p z - r^f z = \varsigma_i\left(\frac{\Pi-P}{P} - r^f z\right)$. Using Equations (52) and (53) and the definition in Equation (56), by the law of total expectation, expected returns read as

$$\begin{aligned}
\mathbb{E}\left(r_i^{pe}z\right) &= \mathbb{E}\left[\mathbb{E}\left(\varsigma_i\left(\frac{\Pi-P}{P} - r^f z\right)|\mathcal{F}_i\right)\right] = \mathbb{E}\left[\left(\frac{1}{\rho\sigma\sqrt{z}}\lambda_i + o(1)\right)\left(\mathbb{E}_i\left(bz|\mathcal{F}_i\right) + \frac{1}{2}\sigma^2 z - pz - r^f z + o(z)\right)\right] \\
&= \frac{1}{\rho}\mathbb{E}\left(\lambda_i^2\right) + o(z) = \frac{1}{\rho}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o(z)
\end{aligned}$$

and the return variance is given by

$$\begin{aligned}
\mathbb{V}\left(r_i^{pe}z\right) = \mathbb{V}\left(r_i^{p}z\right) &= \mathbb{E}\left[\left(r_i^{pe}z\right)^2\right] - \mathbb{E}\left(r_i^{pe}z\right)^2 = \mathbb{E}\left[\left(r_i^{pe}z\right)^2\right] \\
&= \mathbb{E}\left[\varsigma_i^2\mathbb{E}\left(\left(\frac{\Pi-P}{P} - r^f z\right)^2|\mathcal{F}_i\right)\right] = \mathbb{E}\left[\left(\frac{1}{\rho\sigma\sqrt{z}}\lambda_i + o(1)\right)^2\left(\sigma^2 z + o(z)\right)\right] \\
&= \frac{1}{\rho^2}\mathbb{E}\left(\lambda_i^2\right) + o(z) = \frac{1}{\rho^2}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o(z),
\end{aligned}$$

which shows Equations (46) and (47). Equation (48) follows from dividing (46) by the square root of (47). Observe that both $\mathbb{E}\left(r_i^{pe}z\right)$ and $\mathbb{V}\left(r_i^{pe}z\right)$, rise in a_i because $\mathbb{E}\left(\lambda_i^2\right)$ is an increasing function of x_i.

F.3 An Example

Own-return elasticity. Let $\mathbb{E}(\theta) = 0$, $v(x_i) = \kappa x_i$, and $a_0 > a_i^*(\mathcal{I})$. Then, Equation (45) that pins down the demand for information, simplifies to $h_0(\mathcal{I}) + x_i = \sqrt{\frac{a_i}{2\rho\kappa\sigma^2}}$. By Equations (46) and (56)

$$\begin{aligned}
\mathbb{E}\left(r_i^{pe}z\right) &= -\frac{z}{\rho\sigma^2}\frac{1}{h_0(\mathcal{I}) + x_i} + \frac{z}{\rho\sigma^2}\left[\frac{\sigma_\theta^2}{\mathcal{I}^2}p_\xi^2 + \sigma_b^2\left(1-p_\xi\right)^2\right] + o(z) \\
&= -\frac{z}{\rho\sigma}\sqrt{\frac{2\rho\kappa}{a_i}} + \frac{z}{\rho\sigma^2}\left[\frac{\sigma_\theta^2}{\mathcal{I}^2}\left(1 - \frac{\mathcal{N}}{\mathcal{T}\sigma_b^2}\right)^2 + \sigma_b^2\left(\frac{\mathcal{N}}{\mathcal{T}\sigma_b^2}\right)^2\right] + o(z).
\end{aligned}$$

The return function is increasing and concave in a_i: $\frac{d\mathbb{E}\left(r_i^p z\right)}{da_i} = \frac{z}{\sigma}\sqrt{\frac{\kappa}{2\rho a_i^3}} > 0$ and $\frac{d^2\mathbb{E}\left(r_i^p z\right)}{da_i^2} = -\frac{3}{2}\frac{z}{\sigma}\sqrt{\frac{\kappa}{2\rho a_i^5}} < 0$. Consequently, the own-return elasticity elasticity in a given period is

$$\varepsilon_i^{\mathbb{E}(r^p z),a} \equiv \frac{a_i}{\mathbb{E}\left(r_i^p z\right)}\frac{d\mathbb{E}\left(r_i^p z\right)}{da_i} = \frac{1}{\mathbb{E}\left(r_i^p z\right)}\frac{z}{\sigma}\sqrt{\frac{\kappa}{2\rho a_i}} = \frac{\sqrt{\kappa/\left(2\rho\sigma^2 a_i\right)}}{\mathcal{S}\left(a_i,\{a_j\}_{j\in[0,1]}\right)/\rho + r^f}. \tag{57}$$

Since $\frac{\partial\mathcal{S}\left(a_i,\{a_j\}_{j\in[0,1]}\right)}{\partial a_i} > 0$ and $\frac{\partial\sqrt{\kappa/(2\rho\sigma^2 a_i)}}{\partial a_i} < 0$, the own-return elasticity decreases in a_i.

Cross-return elasticity. It is more tedious to derive the cross-return elasticity. I focus on the case, where $\sigma^2 = \sigma_b^2 = \sigma_\theta^2 = 1$. In the following, I show that

$$\begin{aligned}\gamma_{i,i'}^{\mathbb{E}(r^p z),a} &\equiv \frac{a_{i'}}{\mathbb{E}\left(r_i^p z\right)}\frac{d\mathbb{E}\left(r_i^p z\right)}{da_{i'}}\\ &= \frac{a_{i'}}{\mathbb{E}\left(r_i^p z\right)}\left(\frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{T}}\frac{\partial\mathcal{T}_{i'}}{\partial a_{i'}} + \frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{N}}\frac{\partial\mathcal{N}_{i'}}{\partial a_{i'}} + \frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{I}}\frac{\partial\mathcal{I}_{i'}}{\partial a_{i'}}\right) \equiv \frac{1}{\mathbb{E}\left(r_i^p z\right)}\delta_{i'}^{\mathbb{E}(r^p z),a},\end{aligned}$$

where $\delta_{i'}^{\mathbb{E}(r^p z),a}$ is decreasing in $a_{i'}$ and $\delta_{i'}^{\mathbb{E}(r^p z),a} \geq 0$ for $a_{i'} \leq \tilde{a}$. Then, $\delta_{i'}^{\mathbb{E}(r^p z),a} < 0$ for $a_{i'} > \tilde{a}$ trivially follows by the continuity of the return function. Recall the definitions of the aggregate variables $\mathcal{I} \equiv \int_{i'}\mathcal{I}_{i'}di'$, $\mathcal{N} \equiv \int_{i'}\mathcal{N}_{i'}di'$, and $\mathcal{T} \equiv \int_{i'}\mathcal{T}_{i'}di$. For the given parametrization, $\frac{\partial\mathcal{N}_{i'}}{\partial a_{i'}} = \sqrt{\kappa/\left(2a_{i'}\rho\right)}$ and $\frac{\partial\mathcal{T}_{i'}}{\partial a_{i'}} = 1/\rho$,. Use $\mathcal{I} = \mathcal{T} - h_0\left(\mathcal{I}\right)\mathcal{N}$ to show that

$$\frac{\partial\mathcal{I}_{i'}}{\partial a_{i'}} = \frac{\frac{\partial\mathcal{T}_{i'}}{\partial a_{i'}} - h_0\left(\mathcal{I}\right)\frac{\partial\mathcal{N}_{i'}}{\partial a_{i'}}}{1+2\mathcal{I}\mathcal{N}} = \frac{1/\rho - \left(1+\mathcal{I}^2\right)\sqrt{\kappa/\left(2a_{i'}\rho\right)}}{1+2\mathcal{I}\mathcal{N}}.$$

Since

$$\begin{aligned}\frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial p_\xi} &= \frac{2z}{\rho\sigma^2}\left[\frac{\sigma_\theta^2}{\mathcal{I}^2}p_\xi + \sigma_b^2 p_\xi - \sigma_b^2\right] = \frac{2z}{\rho\sigma^2}\frac{\sigma_\theta^2}{\mathcal{I}^2}\left[1 - \frac{\mathcal{N}}{\mathcal{T}\sigma_b^2} + \frac{\mathcal{I}^2}{\sigma_\theta^2}\frac{\mathcal{N}}{\mathcal{T}}\right]\\ &= \frac{2z}{\rho\sigma^2}\frac{\sigma_\theta^2}{\mathcal{I}^2}\left[1 - h_0\left(\mathcal{I}\right)\frac{\mathcal{N}}{\mathcal{T}}\right] = \frac{2z}{\rho\sigma^2}\frac{\sigma_\theta^2}{\mathcal{I}^2}\frac{\mathcal{I}}{\mathcal{T}} = \frac{2z}{\rho\mathcal{I}\mathcal{T}},\end{aligned}$$

$$\frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{N}} = \frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial p_\xi}\frac{\partial p_\xi}{\partial\mathcal{N}} = -\frac{2z}{\rho\mathcal{I}\mathcal{T}^2}$$

and

$$\frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{T}} = \frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial p_\xi}\frac{\partial p_\xi}{\partial\mathcal{T}} = \frac{2z}{\rho\mathcal{I}\mathcal{T}^2}\frac{\mathcal{N}}{\mathcal{T}}.$$

Furthermore, $\frac{\partial\mathbb{E}\left(r_i^p z\right)}{\partial\mathcal{I}} = -\frac{2z}{\rho\mathcal{I}^3}\left(1 - \frac{\mathcal{N}}{\mathcal{T}\sigma_b^2}\right)^2$. Collecting all terms, the cross-return semi-elasticity in a subperiod can be written as

$$
\begin{aligned}
\delta_{i'}^{\mathbb{E}(r^p z),a} &= a_{i'} \frac{2z}{\rho \mathcal{I}} \left(\frac{1}{\mathcal{T}^2} \frac{\mathcal{N}}{\mathcal{T}} \cdot \frac{\partial \mathcal{T}_{i'}}{\partial a_{i'}} - \frac{1}{\mathcal{T}^2} \cdot \frac{\partial \mathcal{N}_{i'}}{\partial a_{i'}} - \frac{1}{\mathcal{I}^2} \frac{\partial \mathcal{I}_{i'}}{\partial a_{i'}} \right) \\
&= a_{i'} \frac{2z}{\rho \mathcal{I}} \left[\left(\frac{1}{\mathcal{T}^2} \frac{\mathcal{N}}{\mathcal{T}} - \frac{(1-\mathcal{N}/\mathcal{T})^2}{\mathcal{I}^2 (1+2\mathcal{I}\mathcal{N})} \right) \cdot \frac{\partial \mathcal{T}_{i'}}{\partial a_{i'}} + \left(\frac{(1-\mathcal{N}/\mathcal{T})^2}{\mathcal{I}^2 (1+2\mathcal{I}\mathcal{N})} \left(1+\mathcal{I}^2\right) - \frac{1}{\mathcal{T}^2} \right) \cdot \frac{\partial \mathcal{N}_{i'}}{\partial a_{i'}} \right] \\
&= \frac{2z a_{i'}}{\rho \mathcal{T}^2 \mathcal{I}^3 (1+2\mathcal{I}\mathcal{N})} \left[\Omega_{\mathcal{T}} \cdot \frac{1}{\rho} + \Omega_{\mathcal{N}} \cdot \sqrt{\kappa / (2 a_{i'} \rho)} \right], \qquad (58)
\end{aligned}
$$

where

$$
\begin{aligned}
\Omega_{\mathcal{T}} &\equiv \mathcal{I}^2 (1+2\mathcal{I}\mathcal{N}) \frac{\mathcal{N}}{\mathcal{T}} - \mathcal{T}^2 (1-\mathcal{N}/\mathcal{T})^2 \\
&= -\mathcal{I}^2 \left[(1+2\mathcal{I}\mathcal{N}) \left(\frac{\mathcal{I}}{\mathcal{T}} + \mathcal{I}^2 \frac{\mathcal{N}}{\mathcal{T}} \right) + \mathcal{I}^2 \mathcal{N}^2 \right] < 0
\end{aligned}
$$

and

$$
\begin{aligned}
\Omega_{\mathcal{N}} &\equiv \mathcal{T}^2 (1-\mathcal{N}/\mathcal{T})^2 \left(1+\mathcal{I}^2\right) - \mathcal{I}^2 (1+2\mathcal{I}\mathcal{N}) \\
&= \mathcal{I}^4 \left[(1+\mathcal{I}\mathcal{N})^2 + \mathcal{N}^2 \right] > 0.
\end{aligned}
$$

This semi-elasticity declines with $a_{i'}$

$$
\frac{\partial \delta_{i'}^{\mathbb{E}(r^p z),a}}{\partial a_{i'}} = \frac{2z}{\rho \mathcal{T}^2 \mathcal{I}^3 (1+2\mathcal{I}\mathcal{N})} \left[\Omega_{\mathcal{T}} \cdot \frac{1}{\rho} + \frac{1}{2} \Omega_{\mathcal{N}} \cdot \sqrt{\kappa / (2 a_{i'} \rho)} \right] < 0.
$$

To show that $\left[\Omega_{\mathcal{T}} \cdot \frac{1}{\rho} + \frac{1}{2} \Omega_{\mathcal{N}} \cdot \sqrt{\kappa / (2 a_{i'} \rho)} \right] < 0$, first, rearrange

$$
4\sqrt{a_{i'} / (2\kappa\rho)} \left[(1+2\mathcal{I}\mathcal{N}) \left(\frac{\mathcal{I}}{\mathcal{T}} + \mathcal{I}^2 \frac{\mathcal{N}}{\mathcal{T}} \right) + \mathcal{I}^2 \mathcal{N}^2 \right] > \mathcal{I}^2 \left[(1+\mathcal{I}\mathcal{N})^2 + \mathcal{N}^2 \right].
$$

Since, $a_{i'} > a_i^* (\mathcal{I}) \;\; \forall i'$, $\sqrt{\frac{a_{i'}}{2\rho\kappa}} > (1+\mathcal{I}^2) \;\; \forall i'$. Therefore,

$$
\begin{aligned}
4\sqrt{\frac{a_{i'}}{2\rho\kappa}} \left[(1+2\mathcal{I}\mathcal{N}) \left(\frac{\mathcal{I}}{\mathcal{T}} + \mathcal{I}^2 \frac{\mathcal{N}}{\mathcal{T}} \right) + \mathcal{I}^2 \mathcal{N}^2 \right] &> 4\left(1+\mathcal{I}^2\right) \left[(1+2\mathcal{I}\mathcal{N}) \left(\frac{\mathcal{I}}{\mathcal{T}} + \mathcal{I}^2 \frac{\mathcal{N}}{\mathcal{T}} \right) + \mathcal{I}^2 \mathcal{N}^2 \right] \\
&= 4 (1+2\mathcal{I}\mathcal{N}) \frac{\mathcal{I}}{\mathcal{T}} + 3\mathcal{I}^2\mathcal{N}^2 + 3\mathcal{I}^2 + 8\mathcal{I}^3\mathcal{N} + 3\mathcal{I}^4\mathcal{N}^2 - 2\mathcal{I}^3\mathcal{N}^2 \\
&\quad + \mathcal{I}^2 \left[(1+\mathcal{I}\mathcal{N})^2 + \mathcal{N}^2 \right] > \mathcal{I}^2 \left[(1+\mathcal{I}\mathcal{N})^2 + \mathcal{N}^2 \right],
\end{aligned}
$$

where the last inequality follows from the fact that $3\mathcal{I}^4\mathcal{N}^2 > 2\mathcal{I}^3\mathcal{N}^2$ for $\mathcal{I} \geq 1$ and $3\mathcal{I}^2\mathcal{N}^2 > 2\mathcal{I}^3\mathcal{N}^2$ for $\mathcal{I} < 1$. Finally, define $\tilde{a} : \Omega_{\mathcal{N}} \cdot \sqrt{\kappa / (2\tilde{a}\rho)} = -\Omega_{\mathcal{T}} \cdot \frac{1}{\rho}$. By the continuity of the cross-return semi-elasticity, $\delta_{i'}^{\mathbb{E}(r^p z),a} \geq 0$ for all $a_{i'} \leq \tilde{a}$ and $\delta_{i'}^{\mathbb{E}(r^p z),a} < 0$ for all $a_{i'} > \tilde{a}$.

F.4 The Financial Market Equilibrium and Nonlinear Taxation

In this section, I shortly demonstrate that the financial market also microfounds scale dependence when there is a nonlinear capital gains tax, $T_k(\cdot)$, instead of a linear wealth tax. Assume that $T_k(0) = T_k''(0) = 0$. For a nonlinear capital gains tax, it does not matter whether or not information costs are deductible from the tax base.

The reasoning is the same as before (Appendix F.2). Again, the repeated financial market interaction is static such that households optimize

$$\max_{x} \mathbb{E}_i \left(\max_{\varsigma} \mathbb{E}_i \left(u \left[a_i \left(1 + r_i^p z \right) - T_k \left(a_i r_i^p z \right) - v \left(x_i \right) z \right] | \mathcal{F}_i \right) \right)$$

in each period. There exists a log-linear rational expectations equilibrium equilibrium in which the price and the optimal investment in the risky asset can be derived

$$log\left(P\right) = pz = \left(p_0 + p_\xi \xi - r^f\right) z + o\left(z\right)$$

and

$$\varsigma_i = \frac{1}{\rho\sigma\sqrt{z}} \frac{1}{1 - T_k'(0)} \lambda_i + o\left(1\right),$$

using the same approximations as before. Similarly, the demand for information and the equilibrium information read as

$$v'\left(x_i\right) = \frac{a_i}{2\rho}\left(1 + T_k'(0)\right) \mathcal{S}'\left(x_i; \mathcal{I}\right)$$

and

$$\mathcal{I} = \frac{1}{\rho\sigma^2} \frac{1}{1 - T_k'(0)} \int_{a_i^*(\mathcal{I})}^{a_1} \frac{a_i x_i \left(a_i; \mathcal{I}\right)}{h_0\left(\mathcal{I}\right) + x_i\left(a_i; \mathcal{I}\right)} dG\left(a_i\right),$$

where $a_i^*\left(\mathcal{I}\right) \equiv 2v'(0)\, \rho\sigma^2 h_0\left(\mathcal{I}\right)^2 / \left(1 + T_k'(0)\right)$ denotes the threshold wealth. The equilibrium is, again, unique.

Taking stock of equilibrium choices, expected returns and the variance of returns are given by

$$\mathbb{E}\left(r_i^{pe} z\right) = \mathbb{E}\left(r_i^p z\right) - r^f z = \frac{1}{\rho} \frac{1}{1 - T_k'(0)} \mathcal{S}\left(a_i, \left\{a_j\right\}_{j\in[0,1]}\right)$$

and

$$\mathbb{V}\left(r_i^{pe} z\right) = \mathbb{V}\left(r_i^p z\right) = \frac{1}{\rho} \frac{1}{1 - T_k'(0)} \mathbb{E}\left(r_i^{pe} z\right),$$

respectively. For $T_k'(0) = 0$, all expressions coincide with those in Appendix F.2.

G A Life-Cycle Economy

In this section, I develop a standard two-period life-cycle framework, as introduced by Farhi and Werning (2010), for studying nonlinear capital taxation when there is scale dependence. Using this framework, I study the nonlinear tax incidence and optimal taxation in partial and in general equilibrium. Moreover, I deal with the presence of other policies. Firstly, I consider a subsidy on the costs of information acquisition (financial advisory). Secondly, I study a financial education program.

G.1 Environment

In the following, I describe the economic environment. The objective is to provide an accessible setting that reveals the main insights about the nonlinear taxation of capital gains. As in Mirrlees (1971), the economy is populated by a continuum of households $i \in [0, 1]$. The only source of heterogeneity is the productivity of labor. Agent i's earnings ability $w_i \in \mathbb{R}_+$ is distributed according to a c.d.f. F and a p.d.f. f. Without loss of generality, one can order household indices such that wages increase in i. Then, one may interpret i as the household's position in the pre-tax wage distribution.

Time is discrete, and there are two periods $t = 0, 1$. In the first period, households supply labor, consume and save. In the second period, they consume their savings. Therefore, the first period may be interpreted as an individual's working life with duration H, whereas, in the second period, she is retired. Individuals may take efforts to increase their returns on investment. The resulting return function increases in the amount of savings. In Section E, I show how this relationship emerges in a financial market setting with optimal portfolio choice and information acquisition. This setting gives rise to inequality in the returns to investment. In the financial market, high-income individuals decide to save more and acquire more information than low-income individuals. This information advantage allows them to generate higher (risk-adjusted) returns than households from lower parts of the income and wealth distribution.

Preferences and technology. Households have Greenwood, Hercowitz, and Huffman (1988) preferences

$$u(l_i, a_i, e_i; w_i) \equiv u_0(c_{i,0} - v_0(l_i)) + \beta u_1(c_{i,1} - v_1(x_i)), \tag{59}$$

where $\beta \in (0, 1]$ denotes the households' discount factor, $u_t(\cdot)$ is a concave and increasing period utility, and $v_t(\cdot)$ denotes the convex and increasing disutility from effort. A household of type i

can transfer resources across periods by saving assets a_i. In the first (working) period, households supply labor $l_i > 0$ and earn after-tax income $y_i - T_l(y_i)$, where $T_l(\cdot)$ denotes the government's nonlinear tax on labor income $y_i \equiv w_i l_i$. To increase the returns on the investment of assets a_i, a household can take effort $x_i > 0$. Assume that the costs of this effort accrue in the second (retirement) period. Capital gains are, for the moment, given by the reduced form relation $\tilde{r}_i\left(e_i, \{e_j\}_{j\in[0,1]}\right) \equiv r_i > 0$ where $r(\cdot)$ is increasing and concave in its first argument. A straightforward interpretation is that households acquire financial knowledge by employing financial advisers to raise the rate of return on their investment. In partial equilibrium, an individual's investment return only depends on her own effort choice, whereas, in general equilibrium, an individual's investment returns may depend on choices by everyone else. In Section E, I microfound this reasoning. Capital gains, $a_{R,i} \equiv r_i a_i$, are taxed nonlinearly according to $T_k(\cdot)$. In Section C, I assume that households can deduct effort costs $v_1(\cdot)$ from the tax base. For completeness, in this section, I consider the situation in which these costs are not deductible. One can show that Lemma 1 in Section E ($\frac{dr_i}{da_i} > 0$) holds in this economy (for $T_k(0) = 0$, $T_k'(0) = 0$, and $T_k''(0) = 0$) irrespective of this deductibility. Let all functions be twice continuously differentiable in their arguments.

Monotonicity. Define the local rate of tax progressivity as $p_t(y) \equiv -\frac{\partial log\left[1-T_t'(y)\right]}{\partial log(y)} = \frac{yT_t''(y)}{1-T_t'(y)}$ for $t \in \{l, k\}$. Observe that the usual monotonicity conditions will hold if labor and capital taxes are not too progressive ($p_l(y_l) < 1$ and $p_k(a_{R,i}) < 1$). That is, effort choices, as well as savings, and, hence, labor and capital income are increasing in the index i. Intuitively, the higher an individual's hourly wage, the more she will work, and the more resources she can transfer to the retirement period. Moreover, an individual's incentives to take efforts to increase her capital gains rise with her position in the pre-tax wage distribution. Due to the one-to-one mapping between wages and incomes, one may write returns as a function of savings, $\tilde{r}_i\left(e_i, \{e_j\}_{j\in[0,1]}\right) = r_i\left(a_i, \{a_j\}_{j\in[0,1]}\right)$. I will make use of this formulation later on.

Household problem. In the working period, households consume their after-tax labor income net of savings

$$c_{i,0} + a_i \le w_i l_i - T_l(w_i l_i). \tag{60}$$

In the retirement period, their consumption is given by their final after-tax wealth

$$c_{i,1} \le a_i(1 + r_i) - T_k(r_i a_i). \tag{61}$$

Let $\mathcal{U}_i(T_l, T_k)$ denote household i's indirect utility from optimally choosing savings, a_i, and effort levels, $\{l_i, x_i\}$, to maximize Equation (59) subject to Equations (60) and (61). As standard, suppose the household problem is convex. With a slight abuse of notation, let l_i and a_i denote household i's Marshallian (uncompensated) labor supply and savings functionals. The first-order conditions of the household maximization problem define these functionals implicitly.

Government problem. For simplicity, suppose that households and the government face the same discount factor. Then, the government's budget constraint reads as

$$\mathcal{R}\left(T_l, T_k\right) \equiv \int_0^1 T_l\left(w_i l_i\right) di + \beta \int_0^1 T_k\left(r_i a_i\right) di \geq \bar{E}. \tag{62}$$

The government has a utilitarian objective function. Consequently, it chooses a tax system $\{T_l, T_k\}$ to maximize

$$\mathcal{G}\left(T_l, T_k\right) \equiv \int_0^1 \varGamma_i \mathcal{U}_i\left(T_l, T_k\right) di \tag{63}$$

subject to Equation (62), where $\varGamma_i$ denotes household i's Pareto weight with $\int_0^1 \varGamma_i di = 1$. Denote λ as the marginal value of public funds and $g_{i,t} \equiv (1/\lambda)\, \varGamma_i u_t'\left(c_{i,t} - v_t\left(\cdot\right)\right)$ as the marginal social welfare weight.

G.2 Incidence of Nonlinear Tax Reforms

In this section, I study the impact of a small reform of an arbitrary (potentially suboptimal) tax scheme, e.g., the US tax code, on labor supply and savings by households, as well as on government revenues and social welfare. Technically, I derive the impact of perturbing an arbitrary tax schedule T_t, where $t \in \{l, k\}$, (e.g., the capital gains tax) on the optimal choices by an agent i and aggregate variables in partial and general equilibrium. In other words, I reform the initial tax schedule by $\hat{T}_t$ and analyze the effects on optimal choices. As a by-product, I obtain the optimal tax scheme when the aggregate marginal benefits are equal to the marginal costs.

Gateaux derivatives. To formalize this idea, define the Gateaux derivative of the functional $\mathscr{F}: \mathcal{C}\left(\mathbb{R}_+, \mathbb{R}\right) \to \mathbb{R}$ at T_t in the direction $\hat{T}_t$ by

$$\hat{\mathscr{F}}\left(T_t, \hat{T}_t\right) \equiv \lim_{\mu \to 0} \frac{d}{d\mu} \mathscr{F}\left(T_t + \mu \hat{T}_t\right).$$

Accordingly, perturb the system of first-order conditions by $\mu \hat{T}_t$ and denote $\hat{l}_i\left(T_t, \hat{T}_t\right)$ and $\hat{a}_i\left(T_t, \hat{T}_t\right)$ the Gateaux derivative of labor supply and savings in the direction $\hat{T}_t$. Similarly, perturb Equations (62) and (63) to obtain the incidence on tax revenues, $\hat{\mathcal{R}}\left(T_t, \hat{T}_t\right)$, and social welfare, $\hat{\mathcal{G}}\left(T_t, \hat{T}_t\right)$.

Elasticities. Denote $I_{i,0} \equiv y_i T_l'\left(y_i\right) - T_l\left(y_i\right)$ and $I_{i,1} \equiv a_{R,i} T_k'\left(a_{R,i}\right) - T_k\left(a_{R,i}\right)$ as the virtual income of individual i in period 0 and period 1, respectively. Define $\zeta_i^{a,\left(1-T_t'\right)}$ $\left(\eta_i^{a,I_t}\right)$ as the compensated elasticity of household i's savings with respect to the retention rate of the tax in period t (the income effect parameter of savings with respect to income in period t) along the nonlinear budget line. The elasticities of labor supply are defined accordingly. Again, let $\tilde{\zeta}$ and $\tilde{\eta}$ indicate the elasticities at a fixed rate of return that, without scale dependence, coincide with the observed elasticities. Given the GHH preferences, labor supply is independent of the capital gains tax scheme ($\zeta_i^{l,\left(1-T_k'\right)} = \eta_i^{l,I_2} = 0$). Moreover, let $\zeta_i^{a,r}$ be the elasticity of savings with respect to the rate of return.[24]

The novelty of this paper to let an individual's rate of return vary with her savings and, in general equilibrium, with the savings of others. As before, define the *own-return elasticity* as $\varepsilon_i^{r,a} \equiv \frac{\partial log[r_i(\cdot)]}{\partial log(a_i)}$. It measures the impact of one's wealth on her rate of return, thus, accounting for scale dependence originating, for example, from the variable acquisition of financial knowledge as in Section E. For all $i' \in [0,1]$ the *cross-return elasticity* $\gamma_{i,i'}^{r,a} \equiv \frac{\partial log[r_i(\cdot)]}{\partial log(a_{i'})}$ captures any kind of complementarity between households' wealth and its return. In the example of Section E, it contains inter-household spillovers from financial knowledge and risk-taking. The cross-return elasticity quantifies in reduced form the impact of the portfolio size of household i' in the returns of household i. In partial equilibrium, it is equal to zero for all i, i'.

Incidence on savings. In the following, I characterize the nonlinear incidence of capital gains tax reforms on savings for a given labor tax. One may write, as an intermediate step, the percentage change of savings in reaction to a capital gains tax reform as

$$\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)}{a_i} = -\underbrace{\tilde{\zeta}_i^{a,\left(1-T_k'\right)}}_{>0}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \underbrace{\tilde{\eta}_i^{a,I_2}}_{\leq 0}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)} + \underbrace{\zeta_i^{a,r}}_{>0}\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)}{r_i}. \tag{64}$$

The first two terms describe the standard positive income and negative substitution effect. As the last term reveals, an inequality multiplier effect from the adjustment in the rate of return, now, augments the reaction of savings.

In the following, I show how to use estimates on the elasticity of returns. The partial equilibrium return adjustment in response to the tax reform is proportional to the reaction of the portfolio size

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{PE}}{r_i} = \varepsilon_i^{r,a}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{PE}}{a_i}. \tag{65}$$

[24]Similar to $\tilde{\zeta}_i^{a,\left(1-T_k'\right)}$ and $\tilde{\eta}_i^{a,I_k}$, the definition of $\zeta_i^{a,r}$ involves a correction factor that accounts for behavioral effects along the nonlinear budget line $\frac{1}{1+p_k(a_{R,i})\tilde{\zeta}_i^{a,(1-\tau_k)}}$, where $\tilde{\zeta}_i^{a,(1-\tau_k)}$ is the compensated savings elasticity along the linear budget line.

In general equilibrium, one needs to account for all kinds of spillovers

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{GE}}{r_i}=\varepsilon_i^{r,a}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{GE}}{a_i}+\int_{i'}\gamma_{i,i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'. \tag{66}$$

Thus, in both cases, one needs to upward adjust income and substitution effects of savings by an inequality multiplier effect $\phi_i\equiv\frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}>1$. As the adjustment in savings depends on the shape of the tax reform, the government can directly redistribute the return inequality.

In general equilibrium, there are also cross-return effects. Therefore, combining Equations (64) and (66), the incidence on savings is given by a Fredholm integral equation of the second kind that can be solved using a standard resolvent formalism. The first lemma characterizes the incidence of a reform of the capital gains tax in closed form.

Lemma 4 (Incidence on savings)**.** *Consider a small reform of an arbitrary capital gains tax scheme in the direction $\hat{T}_k$. Define $\phi_i\equiv\frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}$. In partial equilibrium, the first-order change in the optimal savings is given by*

$$\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{PE}}{a_i}=-\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}-\phi_i\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)}. \tag{67}$$

Let $\int_{i'}\int_i\left|\phi_i\zeta_i^{a,r}\gamma_{i,i'}^{r,a}\right|^2 didi'<1$. Then, the general equilibrium adjustment is given by

$$\begin{aligned}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{GE}}{a_i}=&-\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}-\phi_i\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)}\\&-\phi_i\zeta_i^{a,r}\int_{i'}\phi_{i'}R_{i,i'}\left[\tilde{\zeta}_{i'}^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i'}\right)}{1-T_k'\left(a_{R,i'}\right)}+\tilde{\eta}_{i'}^{a,I_2}\frac{\hat{T}_k\left(a_{R,i'}\right)}{a_{R,i'}\left(1-T_k'\left(a_{R,i'}\right)\right)}\right]di',\end{aligned} \tag{68}$$

where for every $i,i'\in[0,1]$ the resolvent is given by $R_{i,i'}\equiv\sum_{n=1}^{\infty}\mathcal{K}_{i,i'}^{(n)}$ with $\mathcal{K}_{i,i'}^{(1)}=\gamma_{i,i'}^{r,a}$ and, for $n\geq 2$, $\mathcal{K}_{i,i'}^{(n)}=\int_{i''}\mathcal{K}_{i,i''}^{(n-1)}\phi_{i''}\zeta_{i''}^{a,r}\gamma_{i'',i'}^{r,a}di''$.

Proof. Appendix H.2. □

Lemma 4 describes the reaction of savings to a small change in the capital gains tax in terms of sufficient statistics (Chetty (2009)). All these sufficient statistics are, in principle, observable by the econometrician and serve as primitives of the model. Nonetheless, these objects are endogenous variables evaluated at a given tax scheme and equilibrium concept.

Incidence on savings in partial equilibrium. As usual, a change in an individual's capital gains tax induces an income effect and a substitution effect on savings. A rise in the marginal capital gains tax reduces the incentive to transfer resources across periods (substitution effect). At the same time, the household feels poorer in the second period and, therefore, saves more (income effect).

Relative to the case of exogenous capital gains ($\phi_i = 1$), these two effects need to be adjusted upwards by an inequality multiplier effect $\phi_i > 1$ accounting for the endogeneity of returns, which is the main difference of this paper from the existing literature. Tax reforms generate novel inequality multiplier effects. Consider the partial equilibrium economy without income effects and suppose, for instance, that individual i faces a reduction in the marginal capital gains tax. Due to the substitution effect, she will save more. However, the scale dependence leads to an adjustment in her investment returns.

In the financial market example, considered in Section 3, the altered amount of savings triggers the following chain of reactions. Because the individual saves more, she invests a higher absolute amount on the stock market. The larger portfolio raises the incentives to acquire costly information about the fundamentals of the economy that drive the stocks' payoffs. As the individual makes more informed decisions on the financial market, her returns rise. Since her returns on investment become larger relative to before, the payoffs from investment and, therefore, savings increase. The higher amount of savings feeds back into the optimal knowledge acquisition and, in turn, boosts returns. This loop continues infinitely.

The term ϕ_i captures this infinite sequence of adjustments. To see this, rewrite $\phi_i = \frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}} = \sum_{n=0}^{\infty} (\zeta_i^{a,r}\varepsilon_i^{r,a})^n$. Therefore, one can interpret the endogeneity of portfolio returns as an amplification force. It multiplies the standard income and substitution effect. As a result, I establish a version of Proposition 1 (b): savings, just as capital income, react more elastic to reforms of the capital gains tax

$$\phi_i \tilde{\zeta}_i^{a,(1-T_k')} > \tilde{\zeta}_i^{a,(1-T_k')} \quad and \quad \left|\phi_i \tilde{\eta}_i^{a,I_2}\right| \geq \left|\tilde{\eta}_i^{a,I_2}\right|.$$

Incidence on savings in general equilibrium. In general equilibrium, a household's rate of return is a function of everyone's decisions. Therefore, in addition to the described inequality multiplier effects, cross-return effects come into play, which I characterize in closed form in Lemma 4. The additional term aggregates the partial-equilibrium reactions by households across the skill distribution. They are weighted by the resolvent of the integral equation and account for an infinite sequence of return adjustments due to the general equilibrium spillovers. In the financial market example, they come from the endogeneity of stock prices, which aggregate individuals' information acquisition and risk-taking. For instance, a decrease in the tax rate of the rich makes them acquire relatively more financial knowledge and alter their portfolio composition. As a result, the equilibrium price adjusts, which also affects households from the bottom of the wealth distribution, given that they participate in the stock market. However, their altered behavior feeds, in turn, back into the equilibrium price so that the rich modify their choices again.

The resolvent formalism captures this infinite sequence of reactions. The resolvent is the sum of iterated kernels. The first kernel, $\mathcal{K}^{(1)}_{i,i'}$, describes the impact of savings by household i' on the returns of i. The second kernel $\mathcal{K}^{(2)}_{i,i'} = \int_{i''} \gamma^{r,a}_{i,i''} \phi_{i''} \zeta^{a,r}_{i''} \gamma^{r,a}_{i'',i'} di'$ accounts for the effect of savings by i' on the returns and, therefore, savings of households i'' which, in turn, affect the decision making of household i. For $n = 3$ the formula describes the impact of household i' on households i'' who affect i'''. The latter, then, influence the returns generated by household i. Observe that this reasoning is in its spirit similar to Sachs et al. (2020) who study general equilibrium reactions of wages and labor supply in response to a reform of the labor income tax schedule.

Whether or not the presence of general equilibrium adjustments amplifies savings responses depends on the sign and magnitude of $\gamma^{r,a}_{i,i'}$ along the wealth distribution. Suppose, for instance, that $\gamma^{r,a}_{i,i'} > 0$ for all $i' \in [0, 1]$. That is, there is a positive complementarity between a household's return on investment and others' investment. Then, general equilibrium forces further amplify income and substitution effects.

Conversely, suppose households live in a small open economy. Then, they have access to an international financial market, where they interact with other, larger investors or institutions whose decisions are affected by other margins and policies. Thus, the marginal impact of the former households on prices becomes small such that $\gamma^{r,a}_{i,i'} \to 0$.

Incidence on return inequality. One may decompose the incidence on returns in closed form

$$\frac{\hat{r}_i\left(T_k, \hat{T}_k\right)^{GE}}{r_i} = -\phi_i \varepsilon^{r,a}_i \tilde{\zeta}^{a,(1-T'_k)}_i \frac{\hat{T}'_k\left(a_{R,i}\right)}{1 - T'_k\left(a_{R,i}\right)} - \phi_i \varepsilon^{r,a}_i \tilde{\eta}^{a,I_2}_i \frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1 - T'_k\left(a_{R,i}\right)\right)} + \phi_i CE_i, \tag{69}$$

where $CE_i \equiv -\int_{i'} \phi_{i'} R_{i,i'} \left[\tilde{\zeta}^{a,(1-T'_k)}_{i'} \frac{\hat{T}'_k\left(a_{R,i'}\right)}{1-T'_k\left(a_{R,i'}\right)} + \tilde{\eta}^{a,I_2}_{i'} \frac{\hat{T}_k\left(a_{R,i'}\right)}{a_{R,i'}\left(1-T'_k\left(a_{R,i'}\right)\right)}\right] di'$ summarizes the cross effects. Therefore, the return inequality directly depends on the underlying tax code and how the policy-maker wishes to reform it.

One may measure the inequality in returns by their variance, $\mathbb{V}\left(r_i\right)$. The effect of a tax reform on return inequality is, then, given by $\hat{\mathbb{V}}\left(r_i\right) = \hat{\mathbb{E}}\left(r_i^2\right) - \hat{\mathbb{E}}\left(r_i\right)^2$. For simplicity, suppose that there are no income and general equilibrium effects and that the other elasticities are constant along the wealth distribution. Let the tax rate be linear. Then, one can write the impact of a tax reform on the return inequality as

$$d\mathbb{V}(r_i) = -2\mathbb{V}(r_i)(\phi - 1)\frac{\tilde{\zeta}^{a,\left(1-T_k'\right)}}{\zeta^{a,r}}\frac{d\tau_k}{1-\tau_k}.$$

Put differently, the elasticity of return inequality with respect to the capital gains tax

$$\zeta^{\mathbb{V}(r),(1-\tau_k)} \equiv -\frac{\partial log\left[\mathbb{V}(r_i)\right]}{\partial log\left(1-\tau_k\right)} = -\frac{2\varepsilon^{r,a}\zeta^{a,\left(1-T_k'\right)}}{1-\varepsilon^{r,a}\zeta^{a,r}}$$

is negative for $\varepsilon^{r,a} > 0$. Therefore, a rise in the linear tax rate (more redistribution) compresses the distribution of returns and, hence, reduces the inequality in returns. One can also show that mitigating the return inequality goes along with the cost of lowering mean returns. This is, again, Proposition 1 (d).

Incidence on utilities. Having derived the incidence on savings and returns, we can now study the effects on utilities. In partial equilibrium, this simply reads as

$$\hat{\mathcal{U}}_i\left(T_k, \hat{T}_k\right)^{PE} = -\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right) \tag{70}$$

which is a straightforward application of the envelope theorem. In general equilibrium, one needs to keep track of the spillovers, or cross-effects, from others' decisions. That is,

$$\hat{\mathcal{U}}_i\left(T_k, \hat{T}_k\right)^{GE} = -\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right) + \lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\left(1+\zeta_i^{a,r}\right)CE_i. \tag{71}$$

For any equilibrium concept, a rise in the tax liability mechanically reduces the utility of a household. By the envelope theorem, there is no first-order effect due to a change in savings and effort choices. In general equilibrium, due to the endogeneity of portfolio returns, one needs to add the impact of others' decisions on individual investment returns. Not surprisingly, an increase in the rate of return raises the utility of a household. Whether or not returns rise, depends, as described, on the distribution of cross-return elasticities.

Incidence on government revenues and social welfare. Now, one can bring together all parts of the incidence analysis to study the change in social welfare and government revenues in response to the reform of the capital gains tax.

Lemma 5 (Incidence on revenues and welfare)**.** *Let* $\int_{i'}\int_i \left|\phi_i\zeta_i^{a,r}\gamma_{i,i'}^{r,a}\right|^2 didi' < 1$. *Denote* $EQ \in \{PE, GE\}$ *as the equilibrium concept. Then, the first-order change in social welfare in response to a small reform in* T_k *reads as*

$$\hat{\mathcal{G}}\left(T_k, \hat{T}_k\right)^{EQ} = \int_i \Gamma_i\hat{\mathcal{U}}_i\left(T_k, \hat{T}_k\right)^{EQ} di. \tag{72}$$

The first-order change in government revenues is given by

$$\hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{EQ}=\beta\int_i \hat{T}_k\left(a_{R,i}\right)di+\beta\int_i T_k'\left(a_{R,i}\right)a_{R,i}\left[\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{EQ}}{r_i}+\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{EQ}}{a_i}\right]di. \tag{73}$$

Proof. Appendix H.2. □

I start with the impact on revenues. Observe that there are three types of effects: mechanical, behavioral, and return effects. The mechanical and behavioral effects are standard. The first one measures the direct impact of reforming the tax scheme on revenue collection, holding the tax base fixed. The second one regards the change in household behavior in response to a tax reform. In general equilibrium, this adjustment of behavior carries the spillover effects mentioned above. The return on investment adjusts due to changes in an individual's investment size and, in general equilibrium, others' amount of investment.

The effects on welfare are similar. By the envelope theorem, there are no first-order behavioral effects. However, households suffer from a rise in the overall tax liability (mechanical effect). Furthermore, the general equilibrium adjustment of returns imposes uninternalized externalities on individuals. In other words, since an individual's rate of return depends on everyone's choices, one needs to add this additional impact on individual utilities from the behavior of others. In the aggregate, these effects add to the standard mechanical effect on social welfare.

G.3 Optimal Nonlinear Taxation

In this section, I describe the optimal nonlinear capital gains tax for a given labor tax. This procedure is similar to Section C, where I explicitly address the interdependence of labor and capital taxes.

Having studied the nonlinear incidence of arbitrary capital gains tax reforms on government revenues and social welfare, I obtain, as a special case, the optimal capital gains tax by equating the sum of first-order effects equal to zero (see, for example, Saez (2001)). At the optimal tax scheme, there are no first-order effects from reforming the tax scheme in the direction of $\hat{T}_k$:

$$\frac{1}{\lambda}\hat{\mathcal{G}}\left(T_k,\hat{T}_k\right)^{EQ}+\hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{EQ}=0. \tag{74}$$

In other words, one cannot find a revenue-neutral tax reform that raises social welfare. Denote $h\left(a_{R,i}\right)$ as the pdf and $H\left(a_{R,i}\right)$ as the cdf of capital income $a_{R,i}$.

Optimal taxation in partial equilibrium. As a benchmark, I consider the optimal nonlinear tax scheme in partial equilibrium. That is, let $\gamma_{i,i'}^{r,a} = 0$ for all $i, i' \in [0,1]$. Proposition 4 characterizes the optimal nonlinear capital gains tax.

Proposition 4 (Optimal nonlinear capital gains tax in partial equilibrium)**.** *The optimal nonlinear capital gains tax on capital gains in partial equilibrium is almost everywhere given by*

$$\frac{T_k'\left(a_{R,i}\right)^{PE}}{1-T_k'\left(a_{R,i}\right)^{PE}} = \frac{1}{\zeta_i^{a_R,\left(1-T_k'\right)}} \frac{1-H\left(a_{R,i}\right)}{a_{R,i}h\left(a_{R,i}\right)} \int_{a_{R,i}}^{\infty} \left(1-g_{i'',1}\right) exp\left[-\int_{a_{R,i}}^{a_{R,i'''}} \frac{\eta_{i'}^{a_R,T_k}}{\zeta_{i'}^{a_R,\left(1-T_k'\right)}} \frac{da_{R,i'}}{a_{R,i'}}\right] \frac{dH\left(a_{R,i''}\right)}{1-H\left(a_{R,i}\right)}, \tag{75}$$

where $\zeta_i^{a_R,\left(1-T_k'\right)} \equiv \Phi_i \tilde{\zeta}_i^{a,\left(1-T_k'\right)}$, $\eta_i^{a_R,T_k} \equiv \Phi_i \tilde{\eta}_i^{a,T_k}$, $\Phi_i \equiv \left(1+\varepsilon_i^{r,a}\right)\phi_i$, *and* $\phi_i \equiv \frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}$.

Proof. Appendix H.3. ☐

The optimal marginal tax rate on capital gains in partial equilibrium is a version of the Diamond (1998) ABC-formula with income effects (as in Saez (2001)). It expresses the optimal tax wedge on capital gains in terms of behavioral and income effects, the hazard ration of the capital gains distribution, and the marginal social welfare weights above $a_{R,i}$.

Therefore, I obtain Proposition 1 (a). Whether or not rates of return are endogenous, the optimal capital gains tax is described by the observed income and behavioral effects and the observed capital income distribution. Nonetheless, the formation of rates of return directly affects these sufficient statistics.

Observe that Φ_i upwards adjusts the elasticities. Holding the elasticities $\tilde{\zeta}_i^{a,\left(1-T_k'\right)}$ and $\tilde{\eta}_i^{a,T_k}$ fixed, under scale dependence ($\varepsilon_i^{r,a} > 0$ and $\Phi_i > 1$), the compensated elasticity of capital gains is larger

$$\zeta_i^{a_R,\left(1-T_k'\right)} > \tilde{\zeta}_i^{a_R,\left(1-T_k'\right)} = \tilde{\zeta}_i^{a,\left(1-T_k'\right)}$$

than under type dependence only. The income effect does not alter the optimal capital tax since

$$\zeta_i^{a_R,\left(1-T_k'\right)} / \eta_i^{a_R,T_k} = \tilde{\zeta}_i^{a,\left(1-T_k'\right)} / \tilde{\eta}_i^{a,T_k}.$$

Accordingly, the adjustment of elasticities provides a force for a lower capital gains tax under scale dependence. Simultaneously, scale dependence may increase the observed capital income inequality measured by the hazard ratio $\frac{1-H\left(a_{R,i}\right)}{a_{R,i}h\left(a_{R,i}\right)}$, which calls for higher taxes. To establish Part

(c) of Proposition 1, I express Equation (75) in terms of primitives:

$$\frac{T_k'\left(a_R\left(w_i\right)\right)^{PE}}{1-T_k'\left(a_R\left(w_i\right)\right)^{PE}} = \frac{\tilde{\zeta}_i^{a,\left(1-T_l'\right)}}{\zeta_i^{a,\left(1-T_k'\right)}} \frac{1-F\left(w_i\right)}{w_i f\left(w_i\right)} \int_{w_i}^{w_1} \left(1-g_{w_{i''},1}\right) exp\left[-\int_{w_i}^{w_{i''}} \frac{\tilde{\eta}_{w_{i'}}^{a,T_k}}{\zeta_{w_{i'}}^{a,\left(1-T_k'\right)}} \frac{dw_{i'}}{w_{i'}}\right] \frac{dF\left(w_{i''}\right)}{1-F\left(w_i\right)}. \tag{76}$$

Therefore, when investment rates are endogenously determined (scale dependence), the capital gains tax is ceteris paribus the same as the one with exogenous, type-dependent returns on investment holding all the primitives of the economy fixed. The upward adjustment in the savings elasticity just offsets the rise in observed inequality.

To further illustrate the implications for redistribution, suppose the capital gains tax is approximately linear “at the top”, e.g., for the top 1% in the wealth distribution. Assume that the elasticities converge to the values $\zeta^{a_R,\left(1-T_k'\right)} = \varPhi\tilde{\zeta}^{a,\left(1-T_k'\right)}$ and that there are no income effects $\tilde{\eta} = 0$. Suppose that, under type dependence, capital gains in this top bracket are Pareto distributed with parameter $\tilde{a}_k > 1$. Under scale dependence, the Pareto parameter is given by $a_k = \tilde{a}_k/\varPhi$. Then, the linear top tax rate reads as

$$\tau_k^{top} = \frac{1-\bar{g}_k}{1-\bar{g}_k + a_k\zeta^{a_R,\left(1-T_k'\right)}}$$

where $\bar{g}_k$ is the limiting value of the social welfare weight. Therefore, I also obtain the neutrality result at the top: This rise in capital income inequality that scale dependence triggers and the adjustment in the elasticity cancel out ($a_k\zeta^{a_R,\left(1-T_k'\right)} = \tilde{\zeta}^{a,\left(1-T_k'\right)}\tilde{a}_k$). This neutrality result provides a potential justification for why capital gains taxes (e.g., in the US) have not increased despite the drastic rise in capital income inequality.

Whereas the Pareto parameter of the capital income distribution is (relatively) easy to observe, the degree to which portfolio returns are scale-dependent and, accordingly, the size of the correct adjustment of the capital gains elasticity is not. Therefore, I demonstrate the effects of observing the elasticity incorrectly given that scale dependence shapes the capital income distribution. I consider the revenue-maximizing capital tax at the top in the following simple calibration. Note that one can reinterpret this tax as the upper bound on the set of Pareto-efficient tax rates (see Werning (2007)). Following Saez and Stantcheva (2018), let $a_k = 1.4$. Set the elasticity of savings with respect to the rate of return equal to 0.5.

In the Left Panel of Figure 10, I compare the adjusted and unadjusted capital gains elasticities for different combinations of $\left(\varepsilon^{r,a}, \tilde{\zeta}^{a,\left(1-T_k'\right)}\right)$. The red line indicates the capital gains elasticity as

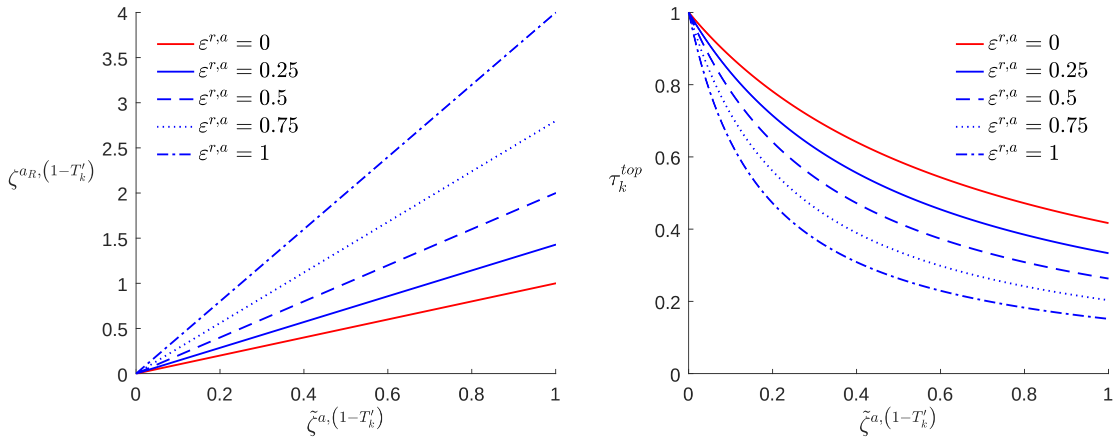

Figure 10: Left Panel: Uncorrected (red) and Corrected (blue) Capital Income Elasticities; Right Panel: Rawlsian Tax Rate at the Top with Uncorrected (red) and Corrected (blue) Elasticities ($a_k = 1.4$, $\zeta^{a,r} = 0.5$, $\eta^{a,I_2} = 0$)

a function of $\tilde{\zeta}^{a,(1-T_k')}$ without incorporating scale dependence, $\varepsilon^{r,a} = 0$ (45 degree line). The blue lines depict the correctly adjusted capital gains elasticities under scale dependence as a function of $\tilde{\zeta}^{a,(1-T_k')}$ for different values of $\varepsilon^{r,a}$. Two things are noteworthy. Firstly, the larger the scale dependence, the greater is the difference between the blue and the red line. Similarly, the difference between the adjusted and the unadjusted elasticity is increasing in the value of $\tilde{\zeta}^{a,(1-T_k')}$. Therefore, the higher is the underlying savings elasticity at a given return or the higher the amount of scale dependence, the more the unadjusted elasticity differs from the adjusted one.

Accordingly, the downward adjustment in the optimal capital gains tax at the top is rising in the amount of size dependence and the value of the underlying savings elasticity (Right Panel of Figure 10). The adjustment is sizable. Therefore, there can be a considerable discrepancy between the optimal tax rate and capital tax set by a policymaker, who ignores the presence of scale dependence and mistakenly assumes that only type dependence generates the observed capital income inequality.

Optimal taxation in general equilibrium. In the following, I characterize the optimal revenue-maximizing nonlinear taxation of capital gains in general equilibrium. For simplicity, abstract from income effects. Moreover, let cross-return elasticities be multiplicatively separable, as in the example of the financial market (Section E.2.2). That is, $\gamma_{i,i'}^{r,a} = \frac{1}{r_i}\delta_{i'}^{r,a}$, where $\delta_{i'}^{r,a}$ decreases in i' and $\delta_{i'}^{r,a} > 0$ ($\delta_{i'}^{r,a} < 0$) for small $a_{i'}$ (large $a_{i'}$). Then, Proposition 5 describes the optimal capital tax.

Proposition 5 (Optimal nonlinear capital gains tax in general equilibrium). *The optimal revenue-maximizing nonlinear capital gains tax on capital gains in general equilibrium is almost everywhere given by*

$$\frac{T_k'(a_{R,i})^{GE}}{1-T_k'(a_{R,i})^{GE}} = \frac{1}{\zeta_i^{a_R,(1-T_k')}}\frac{1-H(a_{R,i})}{a_{R,i}h(a_{R,i})} - \frac{\delta_i^{r,a}}{r_i\left(1+\varepsilon_i^{r,a}\right)(1+\Psi)}$$
$$\times \int_{\mathbb{R}_+} \frac{1+\left(1+\zeta_{i'}^{r,a}\right)\zeta_{i'}^{a,r}}{\zeta_{i'}^{a_R,(1-T_k')}}\frac{1-H(a_{R,i'})}{a_{R,i'}h(a_{R,i'})}\frac{a_{i'}\left[1-T_k'(a_{R,i'})^{GE}\right]}{a_i\left[1-T_k'(a_{R,i})^{GE}\right]}dH(a_{R,i'}), \tag{77}$$

where $\zeta_i^{a_R,(1-T_k')} \equiv \Phi_i \tilde{\zeta}_i^{a,(1-T_k')}$, $\Phi_i \equiv \left(1+\varepsilon_i^{r,a}\right)\phi_i$, $\phi_i \equiv \frac{1}{1-\zeta_i^{a,r}\varepsilon_i^{r,a}}$, *and* $\Psi \equiv \int_{\mathbb{R}_+}\frac{1}{1+\varepsilon_{i'}^{r,a}}\frac{1}{r_{i'}}\delta_{i'}^{r,a}dH(a_{R,i'})$.

Proof. Appendix H.3. □

The optimal tax in general equilibrium adds an additional term on the right-hand side to the partial equilibrium tax (Equation (75)). Observe that the second factor of this extra term is positive (for $0 < T_k'(a_{R,i'}) < 1$ for all $a_{R,i'}$). Therefore, the sign of $\frac{-\delta_i^{r,a}}{r_i\left(1+\varepsilon_i^{r,a}\right)(1+\Psi)}$ determines how to adjust the tax rate in general equilibrium. As in Section C, suppose that cross-return elasticities cancel out $\int_i \gamma_{i,i}^{r,a}di = 0$ and let $\varepsilon_i^{r,a}$ be constant such that $\Psi = 0$. Then, the sign of the adjustment depends on the one of $-\delta_i^{r,a}$.

As a benchmark, consider a politician who sets a tax scheme, $\overline{T}_k'(a_{R,i})$, wrongly assuming that there are no general equilibrium effects for a given initial tax code.[25] Then, one can write the general equilibrium tax rate as

$$\frac{T_k'(a_{R,i})^{GE}}{1-T_k'(a_{R,i})^{GE}} = \frac{\overline{T}_k'(a_{R,i})|_{T_k^{GE}}}{1-\overline{T}_k'(a_{R,i})|_{T_k^{GE}}} - \frac{\delta_i^{r,a}}{r_i\left(1+\varepsilon_i^{r,a}\right)(1+\Psi)}$$
$$\times \int_{\mathbb{R}_+}\left[1+\left(1+\varepsilon_{i'}^{r,a}\right)\zeta_{i'}^{a,r}\right]\frac{\overline{T}_k'(a_{R,i})|_{T_k^{GE}}}{1-\overline{T}_k'(a_{R,i})|_{T_k^{GE}}}\frac{a_{i'}\left[1-T_k'(a_{R,i'})^{GE}\right]}{a_i\left[1-T_k'(a_{R,i})^{GE}\right]}dH(a_{R,i'})$$

Therefore, cross-effects provide a force for higher capital taxes at the top ($\delta_i^{r,a} < 0$ for large a_i) and lower taxes at the bottom making the tax code ceteris paribus more progressive than in the self-confirming policy equilibrium (Proposition 1 (e)).

[25]This notion includes the self-confirming policy equilibrium, proposed by Rothschild and Scheuer (2013, 2016), where a planner implements a tax scheme that generates a capital income distribution for which this tax schedule is optimal, $\overline{T}_k'(a_{R,i})|_{\overline{T}_k}$.

G.4 Other Policies

In the following, I study the interaction with other policies. Consider the partial equilibrium. I distinguish two different policies. In the first case, the government reduces κ for everyone, and, in the second one, it provides a minimum level of financial information. In both cases, the government optimally chooses $\mathcal{P}$ to maximize $\mathcal{G}(T_l, T_k)$ subject to $\mathcal{R}(T_l, T_k) \geq \bar{E} + \beta C(\mathcal{P})$ where $\beta C(\mathcal{P})$ is an increasing and convex cost function. The optimal $\mathcal{P}$ is implicitly defined by

$$\frac{d}{d\mathcal{P}}\left[\frac{1}{\lambda}\mathcal{G}(T_l, T_k) + \mathcal{R}(T_l, T_k)\right] = \beta C'(\mathcal{P}). \tag{78}$$

Using the approximations described in Section E, the first-order condition simplifies to

$$\underbrace{\frac{d}{d\mathcal{P}}\int_i (\Gamma_i/\lambda)\left[u_0(\cdot) + \beta H u_1[\mathbb{E}(\cdot)] + \frac{1}{2}\beta H u_1''(a_i)\mathbb{V}(a_i r_i^p z)\right]di + o(z)}_{\equiv \mathcal{WE}_{\mathcal{P}}}$$
$$+ \underbrace{\frac{d}{d\mathcal{P}}\int_i \beta H T_k[a_i \mathbb{E}(r_i^p z)]\,di + o(z)}_{\equiv \mathcal{RE}_{\mathcal{P}}} = \beta C'(\mathcal{P}). \tag{79}$$

The optimal policy, therefore, trades off first-order revenue and welfare effects. In the following, I describe the first-order condition for each policy in more detail.

Cost subsidy. In the first case, the government lowers the marginal costs of all investors ($\mathcal{P} = \Delta\kappa < 0$). This policy could take the form of a subsidy on financial advisory costs. By the envelope theorem, the first-order welfare impact reduces to the positive effect of cost savings

$$\mathcal{WE}_\kappa = \frac{1}{\kappa}\beta H \int_i g_{i,1}(\mathbb{E}(\cdot))\, x_i z di, \tag{80}$$

whereas the revenue differential includes behavioral effects

$$\begin{aligned}\mathcal{RE}_\kappa = &\frac{1}{\kappa}\beta H \int_i \frac{T_k'[\mathbb{E}(a_i r_i^p z)]}{1 - T_k'[\mathbb{E}(a_i r_i^p z)]}\left(1 + \varepsilon_i^{r,a}\right)\eta_i^{a,I_2} v(x_i)\, z di \\ &- \frac{1}{\kappa}\beta H \int_i T_k'[\mathbb{E}(a_i r_i^p z)]\, a_i \mathbb{E}(r_i^p z)\left(1 + \zeta_i^{a,r}\right)\zeta_i^{\mathbb{E}(r^p z),\kappa} di\end{aligned} \tag{81}$$

with an income effect $\eta_i^{a,I_2} \leq 0$ and the elasticity of the return rate with respect to marginal information costs $\zeta_i^{\mathbb{E}(r^p z),\kappa} \equiv \frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log(\kappa)} < 0$.

On the one hand, the reduction in κ induces a positive impact on capital incomes. Since the acquisition of information becomes relatively cheaper, households acquire more financial knowledge, which allows them to generate higher rates of return. As returns rise, households also save more.

On the other hand, the first term characterizes a negative income effect. Households feel wealthier due to the decline in information costs. As a result, they save less such that capital incomes diminish.

Financial education. In the second case, the government provides a minimum level of financial knowledge as a public good ($\mathcal{P} = \underline{x}$). This policy refers to a situation where the government offers a compulsory finance course to all high school students for free. Formally, the government ensures that $x_i \geq \underline{x}$ for all $i \in [0,1]$. Then, the costs of information acquisition read as $v(x_i) = \kappa z \cdot max\{0, x_i - \underline{x}\}$. Observe that there is a threshold household, $\underline{i}$, with wealth level, $a_{\underline{i}}$, below which households only rely on the education program. They do not acquire any private information beyond $\underline{x}$ and obtain the same rate of return $\mathbb{E}\left(r_{\underline{i}}^p z\right)$. Households above $\underline{i}$ are not affected in their decision making.

The first-order welfare change features two effects

$$\mathcal{WE}_{\underline{x}} = \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \int_0^{\underline{i}} \mathbb{E}\left[g_{i,1}(\cdot)\right] \frac{dlog\left[\mathbb{E}\left(u_1'(\cdot)\right)\right]}{dlog\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]} di + \frac{1}{\underline{x}}\beta H v(\underline{x}) z \int_{\underline{i}}^1 g_{i,1}\left[\mathbb{E}(\cdot)\right] di + o(z) \tag{82}$$

with $\zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \equiv \frac{\partial log\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]}{\partial log(\underline{x})} > 0$. The first one describes the positive impact on utility for households below $\underline{i}$ who experience a rise in their rate of return as the government increases $\underline{x}$ ($d\underline{x} > 0$). The second effect is a mechanical cost-saving effect on households above $\underline{i}$.

The revenue effect

$$\begin{aligned}\mathcal{RE}_{\underline{x}} = & \frac{1}{\underline{x}}\beta H \int_0^{\underline{i}} T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1 + \zeta_i^{a,r}\right) \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} di \\ & + \frac{1}{\underline{x}}\beta H \int_{\underline{i}} \frac{T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]}{1 - T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]} a_i \mathbb{E}\left(r_i^p z\right)\left(1 + \varepsilon_i^{r,a}\right) \eta_i^{a,I_2} v(\underline{x}) z di + o(z)\end{aligned} \tag{83}$$

characterizes income effects for all households and the effects of $\underline{x}$ on the capital incomes of households below $\underline{i}$. A rise in the minimum level of financial knowledge allows these households to obtain higher rates of return. Moreover, they save more as returns increase.

Which of the two policies the government should undertake, depends on the magnitude of the revenue and welfare effects. In particular, one needs to know about the size of the policy elasticities $\zeta_i^{\mathbb{E}(r^p z),\kappa}$ and $\zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}}$. These describe the responsiveness of individual returns with respect to a reduction in information costs and a rise in the minimum education provision by the government, respectively.

The impact of the policy also interacts with the tax code. Two identical societies that only vary in their redistributive preference may, therefore, deem very distinct policies desirable. Similarly, this is the case when they solely differ in the way how returns are formed (i.e., the importance of scale dependence relative to type dependence). Moreover, the marginal costs of policy implementation, $C'(\mathcal{P})$, depend on the respective policy $\mathcal{P}$ and other parameters, such as the efficiency of a country's educational system.

H Proofs of Section G

H.1 Preliminaries

Household choices. For the specified GHH preferences, the households' first-order conditions are given by

$$\begin{aligned} [l_i]: \; 0 &= w_i \left(1 - T_l'\left(w_i l_i\right)\right) - v_0'\left(l_i\right) \\ [a_i]: \; 0 &= \left[1 + r_i\left(1 - T_k'\left(r_i a_i\right)\right)\right] \beta u_1'(\cdot) - u_0'(\cdot) \\ [e_i]: \; 0 &= a_i \left(1 - T_k'\left(r_i a_i\right)\right) r_i'\left(e_i\right) - v_1'\left(e_i\right), \end{aligned} \tag{84}$$

where the optimal labor supply decisions can be decoupled from the savings and information effort choices. Let the second-order conditions hold. That is,

$$\frac{\partial^2 u\left(l_i, a_i, e_i; w_i\right)}{\partial l_i^2} = -w_i^2 T_l''\left(w_i l_i\right) - v_0''\left(l_i\right) < 0$$

and the Hessian $H = \begin{pmatrix} \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i^2} & \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i \partial e_i} \\ \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i \partial e_i} & \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial e_i^2} \end{pmatrix}$ is negative definite.

Monotonicity. Now, I describe the relationship between optimal choices and pre-tax wages. A household's labor supply increases with the wage rate

$$\begin{aligned} \frac{dl_i}{dw_i} &= -\frac{\partial^2 u\left(l_i, a_i, e_i; w_i\right) / \left(\partial l_i \partial w_i\right)}{\partial^2 u\left(l_i, a_i, e_i; w_i\right) / \partial l_i^2} = \frac{1 - T_l'\left(w_i l_i\right) - w_i l_i T_l''\left(w_i l_i\right)}{w_i^2 T_l''\left(w_i l_i\right) + v_0''\left(l_i\right)} \\ &= \frac{l_i}{w_i} \frac{1 - p_l\left(y_i\right)}{\frac{v_0''\left(l_i\right)}{v_0'\left(l_i\right)} l_i + p_l\left(y_i\right)} > 0, \end{aligned}$$

where I use the definition of the local rate of tax progressivity $p_t(y) \equiv \frac{y T_t''(y)}{1 - T_t'(y)}$ for $t \in \{l, k\}$ and the assumption that $p_l(y_i) < 1$. Since $\frac{dy_i}{dw_i} = w_i \frac{dl_i}{dw_i} + l_i$, labor earnings also rise with the wage rate.

Savings and effort choices depend on w_i according to

$$\begin{pmatrix} da_i/dw_i \\ de_i/dw_i \end{pmatrix} = -H^{-1} \begin{pmatrix} -u_0''(\cdot)\, l_i \left(1 - T_l'(w_i l_i)\right) \\ 0 \end{pmatrix}$$

$$= \frac{1}{det(H)} \begin{pmatrix} \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial e_i^2} & -\frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i \partial e_i} \\ -\frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i \partial e_i} & \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i^2} \end{pmatrix} \begin{pmatrix} u_0''(\cdot)\, l_i \left(1 - T_l'(w_i l_i)\right) \\ 0 \end{pmatrix}$$

$$= \frac{u_0''(\cdot)\, l_i \left(1 - T_l'(w_i l_i)\right)}{det(H)} \begin{pmatrix} \frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial e_i^2} \\ -\frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial a_i \partial e_i} \end{pmatrix}.$$

Observe that by the second-order conditions $det(H) > 0$ and $\frac{\partial^2 u(l_i,a_i,e_i;w_i)}{\partial e_i^2} < 0$. Moreover, for $p_k(r_i a_i) < 1$,

$$\frac{\partial^2 u(l_i, a_i, e_i; w_i)}{\partial a_i \partial e_i} = \beta \left(1 - T_k'(a_{R,i})\right) \left(1 - p_k(a_{R,i})\right) r_i'(e_i) > 0.$$

Altogether, due to the concavity of $u_0(\cdot)$, $\frac{da_i}{dw_i} > 0$ and $\frac{de_i}{dw_i} > 0$. Consequently, capital income rises in the pre-tax wage $\frac{da_{R,i}}{dw_i} = a_i r'(e_i) \frac{de_i}{dw_i} + r_i \frac{da_i}{dw_i} > 0$.

H.2 Incidence of Nonlinear Tax Reforms

Incidence on savings in partial equilibrium. To derive the incidence on savings in partial equilibrium, plug Equation (65) into (64)

$$\frac{\hat{a}_i\left(T_k, \hat{T}_k\right)^{PE}}{a_i} = -\tilde{\zeta}_i^{a,(1-T_k')} \frac{\hat{T}_k'(a_{R,i})}{1 - T_k'(a_{R,i})} - \tilde{\eta}_i^{a,I_2} \frac{\hat{T}_k(a_{R,i})}{a_{R,i}\left(1 - T_k'(a_{R,i})\right)} + \zeta_i^{a,r} \varepsilon_i^{r,a} \frac{\hat{a}_i\left(T_k, \hat{T}_k\right)^{PE}}{a_i}$$

Rearrange this expression to obtain Equation (67) in Lemma 4.

Incidence on savings in general equilibrium. To derive Equation (68) in Lemma 4, insert Equation (66) into (64) and rearrange

$$\frac{\hat{a}_i\left(T_k, \hat{T}_k\right)^{GE}}{a_i} = -\phi_i \tilde{\zeta}_i^{a,(1-T_k')} \frac{\hat{T}_k'(a_{R,i})}{1 - T_k'(a_{R,i})} - \phi_i \tilde{\eta}_i^{a,I_2} \frac{\hat{T}_k(a_{R,i})}{a_{R,i}\left(1 - T_k'(a_{R,i})\right)} + \phi_i \zeta_i^{a,r} \int_{i'} \gamma_{i,i'}^{r,a} \frac{\hat{a}_{i'}\left(T_k, \hat{T}_k\right)^{GE}}{a_{i'}} di'$$

This expression is a Fredholm integral equation of the second kind. Suppose that $\int_{i'} \int_i \left|\phi_i \zeta_i^{a,r} \gamma_{i,i'}^{r,a}\right|^2 di di' < 1$. Then, by Theorem 2.3.1 in Zemyan (2012), the unique solution to this expression is given by Equation (68).

Incidence on return inequality. In partial equilibrium, the effect on returns can be written as

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{PE}}{r_i}=-\phi_i\varepsilon_i^{r,a}\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}-\phi_i\varepsilon_i^{r,a}\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)},$$

where I use Equations (65) and (67). Using the fact that

$$\int_{i'}\gamma_{i,i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'=-\int_{i'}\phi_{i'}R_{i,i'}\left[\tilde{\zeta}_{i'}^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i'}\right)}{1-T_k'\left(a_{R,i'}\right)}+\tilde{\eta}_{i'}^{a,I_2}\frac{\hat{T}_k\left(a_{R,i'}\right)}{a_{R,i'}\left(1-T_k'\left(a_{R,i'}\right)\right)}\right]di'\equiv CE_i,$$

the general equilibrium incidence on returns reads as

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{GE}}{r_i}=-\phi_i\varepsilon_i^{r,a}\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}-\phi_i\varepsilon_i^{r,a}\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)}+\phi_i CE_i$$

Incidence on utilities. The partial equilibrium incidence on household utilities is standard. In general equilibrium, one needs to account for cross-return effects that come from the dependence of each household's return rate on the savings of all other households

$$\begin{aligned}\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right)^{GE}=&-\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right)+\lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\int_{i'}\gamma_{i,i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'\\&+\lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\zeta_i^{a,r}\int_{i'}\gamma_{i,i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'\\=&-\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right)+\lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\left(1+\zeta_i^{a,r}\right)CE_i\end{aligned}$$

Incidence on revenues and welfare. Equation (72) is standard. Perturb Equation (62)

$$\hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{EQ}=\beta\int_i\hat{T}_k\left(a_{R,i}\right)di+\beta\int_i T_k'\left(a_{R,i}\right)\left[a_i\hat{r}_i\left(T_k,\hat{T}_k\right)^{EQ}+r_i\hat{a}_i\left(T_k,\hat{T}_k\right)^{EQ}\right]di$$

and rearrange to get (73).

H.3 Optimal Nonlinear Taxation

Optimal taxation in partial equilibrium. Setting the sum of first-order welfare and revenue effects equal to zero, the optimal nonlinear capital gains tax is characterized by

$$\int_{a_{R,i}}\left[1-g_{i,1}-\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\eta}_i^{a,I_2}\frac{T_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}\right]\hat{T}_k\left(a_{R,i}\right)dH\left(a_{R,i}\right)$$
$$=\int_{a_{R,i}}a_{R,i}\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{T_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}\hat{T}_k'\left(a_{R,i}\right)dH\left(a_{R,i}\right).$$

Integrate the first term by parts and apply the fundamental theorem of calculus of variations to get

$$\frac{T_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}=\frac{1}{\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-T_k')}}\frac{1-H\left(a_{R,i}\right)}{a_{R,i}h\left(a_{R,i}\right)}\int_{a_{R,i}}^{\infty}\left[1-g_{i',1}-\left(1+\varepsilon_{i'}^{r,a}\right)\phi_{i'}\tilde{\eta}_{i'}^{a,I_2}\frac{T_k'\left(a_{R,i'}\right)}{1-T_k'\left(a_{R,i'}\right)}\right]\frac{dH\left(a_{R,i'}\right)}{1-H\left(a_{R,i}\right)}.$$

This expression is a first-order linear differential equation. Use standard techniques (see Saez (2001)) to obtain Equation (75).

To express (75) in terms of the pre-tax wage distribution, change the variables in the integration

$$\frac{T_k'\left(a_R\left(w_i\right)\right)}{1-T_k'\left(a_R\left(w_i\right)\right)}=\frac{1}{\zeta_i^{a_R,(1-T_k')}}\frac{1-F\left(w_i\right)}{a_{R,i}h\left(a_{R,i}\right)}\int_{w_i}^{w_1}\left(1-g_{w_{i''},1}\right)exp\left[-\int_{w_i}^{w_{i''}}\frac{\tilde{\eta}_{w_{i'}}^{a,T_k}}{\zeta_{w_{i'}}^{a,(1-T_k')}}\frac{dw_{i'}}{w_{i'}}\right]\frac{dF\left(w_{i''}\right)}{1-F\left(w_i\right)}. \tag{85}$$

Since $F\left(w_i\right)=H\left(a_R\left(w_i\right)\right)$,

$$f\left(w_i\right)w_i=a_R\left(w_i\right)h\left(a_R\left(w_i\right)\right)\frac{da_R\left(w_i\right)}{dw_i}\frac{w_i}{a_R\left(w_i\right)}. \tag{86}$$

The elasticity of capital income with respect to the wage rate is given by

$$\frac{da_R\left(w_i\right)}{dw_i}\frac{w_i}{a_R\left(w_i\right)}=\left(1+\varepsilon_i^{r,a}\right)\frac{da}{dw_i}\frac{w_i}{a}=\left(1+\varepsilon_i^{r,a}\right)\frac{da}{d\left(1-T_l'\left(w_il_i\right)\right)}\frac{1-T_l'\left(w_il_i\right)}{a}=\left(1+\varepsilon_i^{r,a}\right)\tilde{\zeta}_i^{a,(1-T_l')}$$
$$+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\varepsilon_i^{r,a}\frac{da\left(w_i\right)}{d\left(1-T_l'\left(w_il_i\right)\right)}\frac{1-T_l'\left(w_il_i\right)}{a\left(w_i\right)}=\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-T_l')} \tag{87}$$

where the second and fourth equality follow from the fact that

$$\frac{da}{d\left(1-T_l'\left(w_il_i\right)\right)}\frac{1-T_l'\left(w_il_i\right)}{a}=\frac{u_0''\left(\cdot\right)w_il_i\left(1-T_l'\left(w_il_i\right)\right)}{a_idet\left(H\right)}\frac{\partial^2u\left(l_i,a_i,e_i;w_i\right)}{\partial e_i^2}=\frac{da}{dw_i}\frac{w_i}{a}.$$

Plug Equations (86) and (87) into (85), to get Equation (76).

Optimal taxation in general equilibrium. First, note that, in the absence of income effects and for $\gamma_{i,i'}^{r,a} = \frac{1}{r_i}\delta_{i'}^{r,a}$, the incidence on savings can be written as

$$\begin{aligned}
\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)}{a_i} &= -\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} + \zeta_i^{a,r}\frac{1}{r_i}\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)}{a_{i'}}di' \\
&= -\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \zeta_i^{a,r}\frac{1}{r_i}\frac{1}{1-\int_i\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}di}\int_i\delta_i^{r,a}\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}di,
\end{aligned}$$

noting that

$$\begin{aligned}
\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)}{a_{i'}}di' &= -\int_i\delta_i^{r,a}\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}di + \int_i\delta_i^{r,a}\zeta_i^{a,r}\frac{1}{r_i}di\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)}{a_{i'}}di' \\
&= -\frac{1}{1-\int_i\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}di}\int_i\delta_i^{r,a}\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}di.
\end{aligned}$$

By the latter expression, the response of capital income reads as

$$\begin{aligned}
\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)}{a_i} + \frac{\hat{r}_i\left(T_k,\hat{T}_k\right)}{r_i} &= \left(1+\varepsilon_i^{r,a}\right)\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)}{a_i} + \frac{1}{r_i}\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)}{a_{i'}}di' \\
&= -\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]\frac{1}{r_i}\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)}{a_{i'}}di'
\end{aligned}$$

and the incidence on utility is given by

$$\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right) = -\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right) + \lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\frac{1+\zeta_i^{a,r}}{r_i}\int_{i'}\delta_{i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'.$$

Again, impose that there is no first-order effect on the social planner's objective function, $\frac{1}{\lambda}\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right)^{GE} + \hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{GE} = 0$, to characterize the optimal capital gains tax

$$\begin{aligned}
\int_i\left(1-g_{i,1}\right)\hat{T}_k\left(a_{R,i}\right)di &= \int_i g_{i,1}a_{R,i}\left[1-T_k'\left(a_{R,i}\right)\right]\frac{1+\zeta_i^{a,r}}{r_i}di\frac{1}{1-\int_i\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}di}\int_i\delta_i^{r,a}\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}di \\
&+ \int_i T_k'\left(a_{R,i}\right)a_{R,i}\Bigg[\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} \\
&+ \left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]\frac{1}{r_i}\frac{1}{1-\int_{i'}\zeta_{i'}^{a,r}\frac{1}{r_{i'}}\delta_{i'}^{r,a}di'}\int_{i'}\delta_{i'}^{r,a}\phi_{i'}\tilde{\zeta}_{i'}^{a,\left(1-T_k'\right)}\frac{\hat{T}_k'\left(a_{R,i'}\right)}{1-T_k'\left(a_{R,i'}\right)}di'\Bigg]di
\end{aligned}$$

if and only if

$$\int_i (1-g_{i,1})\hat{T}_k(a_{R,i})\,di = \int_i T_k'(a_{R,i})\,a_{R,i}\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'(a_{R,i})}{1-T_k'(a_{R,i})}di$$
$$+\int_i a_i\left\{g_{i,1}\left[1-T_k'(a_{R,i})\right]\left(1+\zeta_i^{a,r}\right)+T_k'(a_{R,i})\left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]\right\}di$$
$$\times\frac{1}{1-\int_i \zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}di}\int_i \delta_i^{r,a}\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'(a_{R,i})}{1-T_k'(a_{R,i})}di.$$

In this setting, the easiest way to derive an expression for the optimal capital gains tax is to consider the Saez (2001) perturbation: $\hat{T}_k(a_{R,i}) = 1_{a_{R,i}\geq a_{R,i^*}}$ and $\hat{T}_k'(a_{R,i}) = \delta_{a_{R,i^*}}(a_{R,i})$, where $\delta_{a_{R,i^*}}(a_{R,i})$ is the Dirac delta function. Then, under revenue maximization ($g_{i,1}=0$), the previous expression simplifies to

$$\frac{T_k'(a_{R,i^*})}{1-T_k'(a_{R,i^*})} = \frac{1}{\left(1+\zeta_{i^*}^{r,a}\right)\phi_{i^*}\tilde{\zeta}_{i^*}^{a,(1-T_k')}}\frac{1-H(a_{R,i^*})}{a_{R,i^*}h(a_{R,i^*})}$$
$$-\frac{\int_{a_{R,i}} a_i T_k'(a_{R,i})\left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]dH(a_{R,i})}{1-\int_{a_{R,i}}\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})}\frac{\frac{1}{r_{i^*}}\delta_{i^*}^{r,a}}{1+\varepsilon_{i^*}^{r,a}}\frac{1}{a_{i^*}\left(1-T_k'(a_{R,i^*})\right)},$$

where I expressed all the variables in terms of observables. Rearrange and integrate out to get

$$\int_{a_{R,i}}\left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]\frac{a_i\left(1-T_k'(a_{R,i})\right)}{\left(1+\varepsilon_{i^*}^{r,a}\right)\phi_{i^*}\tilde{\zeta}_{i^*}^{a,(1-T_k')}}\frac{1-H(a_{R,i})}{a_{R,i}h(a_{R,i})}dH(a_{R,i})$$
$$=\frac{\int_{a_{R,i}}\frac{1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}}{1+\varepsilon_i^{r,a}}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})+1-\int_{a_{R,i}}\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})}{1-\int_{a_{R,i}}\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})}\int_{a_{R,i}} a_i T_k'(a_{R,i})\left[1+\left(1+\zeta_i^{r,a}\right)\zeta_i^{a,r}\right]dH(a_{R,i})$$
$$=\left[1+\int_{a_{R,i}}\frac{1}{1+\varepsilon_i^{r,a}}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})\right]\frac{\int_{a_{R,i}} a_i T_k'(a_{R,i})\left[1+\left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,r}\right]dH(a_{R,i})}{1-\int_{a_{R,i}}\zeta_i^{a,r}\frac{1}{r_i}\delta_i^{r,a}dH(a_{R,i})}.$$

Altogether, one can write the optimal nonlinear capital gains tax as

$$\frac{T_k'(a_{R,i})}{1-T_k'(a_{R,i})} = \frac{1}{\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-T_k')}}\frac{1-H(a_{R,i})}{a_{R,i}h(a_{R,i})}-\frac{\frac{1}{r_i}\delta_i^{r,a}}{1+\varepsilon_i^{r,a}+\int_{a_{R,i}'}\frac{1+\varepsilon_i^{r,a}}{1+\varepsilon_{i'}^{r,a}}\frac{1}{r_{i'}}\delta_{i'}^{r,a}dH(a_{R,i'})}$$
$$\times\int_{a_{R,i'}}\frac{1+\left(1+\varepsilon_{i'}^{r,a}\right)\zeta_{i'}^{a,r}}{\left(1+\varepsilon_{i'}^{r,a}\right)\phi_{i'}\tilde{\zeta}_{i'}^{a,(1-T_k')}}\frac{1-H(a_{R,i'})}{a_{R,i'}h(a_{R,i'})}\frac{a_{i'}\left(1-T_k'(a_{R,i'})\right)}{a_i\left(1-T_k'(a_{R,i})\right)}dH(a_{R,i'}),$$

which concludes the proof of Equation (77).

To compare this capital gains tax to the one in the self-confirming policy equilibrium, note that in the latter case

$$\frac{\overline{T}_k'(a_{R,i})\,|_{T_k^{GE}}}{1-\overline{T}_k'(a_{R,i})\,|_{T_k^{GE}}} = \frac{1}{\zeta_i^{a_R,(1-T_k')}}\frac{1-H(a_{R,i})}{a_{R,i}h(a_{R,i})}$$

and insert this expression into (77).

H.4 Other Policies

Preliminaries. In line with the financial market in Section E, second-period expected utility, which the policy $\mathcal{P}$ affects, is given by

$$H\cdot\mathbb{E}_i\left(u_1\left[a_i\left(1+r_i^p z\right)-T_k\left(a_i r_i^p z\right)-v\left(x_i\right)z\right]\right)=H\cdot u_1\left[a_i\left(1+\mathbb{E}\left(r_i^p z\right)\right)-T_k\left(a_i\mathbb{E}\left(r_i^p z\right)\right)-v\left(x_i\right)z\right]$$
$$+H\cdot\frac{1}{2}u_1''\left(a_i\right)\mathbb{V}\left(a_i r_i^p z\right)+o\left(z\right).$$

Therefore, the impact of policy $\mathcal{P}$ on welfare can be written as

$$\mathcal{WE}_{\mathcal{P}}\equiv\frac{d}{d\mathcal{P}}\frac{1}{\lambda}\mathcal{G}\left(T_l,T_k\right)=\frac{d}{d\mathcal{P}}\int_i\left(\mathit{\Gamma}_i/\lambda\right)\left[u_0\left(\cdot\right)+\beta H u_1\left[\mathbb{E}\left(\cdot\right)\right]+\frac{1}{2}\beta H u_1''\left(a_i\right)\mathbb{V}\left(a_i r_i^p z\right)\right]di+o\left(z\right).$$

A household's tax liability can be approximated by

$$T_k\left(R_i a_i\right)=T_k\left(a_i r_{i,1}z+...+a_i r_{i,H}z\right)+o\left(z\right)\equiv T_k\left(a_i r_{i,1}z,...,a_i r_{i,H}z\right)+o\left(z\right)=\sum_{h=1}^{H}T_k'\left(0\right)a_i r_{i,h}z+o\left(z\right)$$
$$=T_k\left(a_i r_{i,1}z\right)-T_k\left(0\right)+...+T_k\left(a_i r_{i,H}z\right)-T_k\left(0\right)+o\left(z\right)=\sum_{h=1}^{H}T_k\left(a_i r_{i,h}z\right)+o\left(z\right)$$

Using this expression, the first-order effect on revenues reads as

$$\mathcal{RE}_{\mathcal{P}}\equiv\frac{d}{d\mathcal{P}}\mathcal{R}\left(T_l,T_k\right)=\frac{d}{d\mathcal{P}}\int_i\beta\mathbb{E}\left(T_k\left[R_i a_i\right]\right)di=\frac{d}{d\mathcal{P}}\int_i\beta\sum_{h=1}^{H}T_k\left[a_i\mathbb{E}\left(r_{i,h}^p z\right)\right]di+o\left(z\right)$$
$$+\frac{d}{d\mathcal{P}}\int_i\beta\sum_{h=1}^{H}T_k'\left[a_i\mathbb{E}\left(r_{i,h}^p z\right)\right]\mathbb{E}\left[a_i r_{i,h}^p z-a_i\mathbb{E}\left(r_{i,h}^p z\right)\right]di$$
$$+\frac{1}{2}\frac{d}{d\mathcal{P}}\int_i\beta\sum_{h=1}^{H}T_k''\left[a_i\mathbb{E}\left(r_{i,h}^p z\right)\right]\mathbb{E}\left[\left(a_i r_{i,h}^p z-a_i\mathbb{E}\left(r_{i,h}^p z\right)\right)^2\right]di+o\left(z\right)$$
$$=\frac{d}{d\mathcal{P}}\int_i\beta\sum_{h=1}^{H}T_k\left[a_i\mathbb{E}\left(r_{i,h}^p z\right)\right]di+o\left(z\right)=\frac{d}{d\mathcal{P}}\int_i\beta H T_k\left[a_i\mathbb{E}\left(r_i^p z\right)\right]di+o\left(z\right)$$

since, in partial equilibrium,

$$\mathbb{E}\left(r_{i,h}^p z\right)=\frac{1}{\rho}\mathcal{S}\left(a_{i,h}\right)z+r^f z+o\left(z\right)=\frac{1}{\rho}\mathcal{S}\left(a_i\Pi_{s=1}^h\left(1+r_{i,h}^p z\right)\right)z+r^f z+o\left(z\right)$$
$$=\frac{1}{\rho}\mathcal{S}\left(a_i+a_i\sum_{s=1}^{h}r_{i,h}^p z\right)z+r^f z+o\left(z\right)=\frac{1}{\rho}\mathcal{S}\left(a_i\right)z+r^f z+o\left(z\right)=\mathbb{E}\left(r_i^p z\right).$$

Cost subsidy. For a cost subsidy, $\mathcal{P} = \Delta\kappa < 0$, the first-order welfare effect can be written as in Equation (80)

$$\mathcal{WE}_{\kappa} = \int_{i} (\mathit{\Gamma}_i/\lambda)\, \beta H u_1' \left[\mathbb{E}(\cdot)\right] x_i z di \equiv \frac{1}{\kappa}\beta H \int_{i} g_{i,1}\left(\mathbb{E}(\cdot)\right) v\left(x_i\right) z di + o\left(z\right)$$

and, defining the elasticity of returns with respect to marginal information costs $\zeta_i^{\mathbb{E}(r^p z),\kappa} \equiv \frac{\partial log\left[\mathbb{E}\left(r_i^p z\right)\right]}{\partial log(\kappa)} < 0$, the effect on government revenue is given by Equation (81)

$$\begin{aligned}
\mathcal{RE}_{\kappa} = & -\frac{d}{d\kappa}\int_{i}\beta H T_k\left[a_i \mathbb{E}\left(r_i^p z\right)\right] di + o\left(z\right) = -\frac{1}{\kappa}\beta H \int_{i} T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right) \\
& \times \left[\left(1+\varepsilon_i^{r,a}\right)\eta_i^{a,I_2}\frac{-x_i z}{\mathbb{E}\left(r_i^p z\right)\left(1-T_k'\left(a_i\mathbb{E}\left(r_i^p z\right)\right)\right)}\frac{\kappa}{a_i} + \left(1+\zeta_i^{a,r}\right)\zeta_i^{\mathbb{E}(r^p z),\kappa}\right] di + o\left(z\right) \\
& = \frac{1}{\kappa}\beta H\int_{i}\frac{T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]}{1-T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]}\left(1+\varepsilon_i^{r,a}\right)\eta_i^{a,I_2} v\left(x_i\right) z di \\
& - \frac{1}{\kappa}\beta H\int_{i} T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1+\zeta_i^{a,r}\right)\zeta_i^{\mathbb{E}(r^p z),\kappa} di + o\left(z\right).
\end{aligned}$$

Financial education. When the government provides a minimal level of financial knowledge, $\underline{x}$, for free, such that the information cost reads as $v\left(x_i\right) = \kappa z \cdot max\left\{0, x_i - \underline{x}\right\}$, there is a threshold household, below which households do not acquire additional information and obtain the same return rate

$$x_i = \sqrt{\frac{a_i}{\sigma\rho\kappa}} - 1 - \mathcal{I} \leq \underline{x} \iff a_i \leq \sigma\rho\kappa\left(\underline{x}+1+\mathcal{I}\right)^2 \equiv a_{\underline{i}}.$$

Define the elasticity of returns with respect to the minimal information provided by the government as $\zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \equiv \frac{dlog\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]}{dlog(\underline{x})} > 0$. The effect of raising $\underline{x}$ ($d\underline{x} > 0$) on welfare consists of a rise in return rates of households below $\underline{i}$ and a cost reduction for households above $\underline{i}$

$$\begin{aligned}
\mathcal{WE}_{\underline{x}} = & \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}}\int_0^{\underline{i}}\left(\mathit{\Gamma}_i/\lambda\right)\left[u_1'\left[\mathbb{E}(\cdot)\right] a_i\left(1-T_k'\left[\mathbb{E}\left(a_i r_{\underline{i}}^p z\right)\right]\right) + \frac{1}{2}u_1''\left(a_i\right) a_i^2\right]\mathbb{E}\left(r_{\underline{i}}^p z\right) di + o\left(z\right) \\
& + \frac{1}{\underline{x}}\beta H v\left(\underline{x}\right) z\int_{\underline{i}}^1\left(\mathit{\Gamma}_i/\lambda\right) u_1'\left[a_i\left(1+\mathbb{E}\left(r_i^p z\right)\right) - T_k\left(a_i\mathbb{E}\left(r_i^p z\right)\right) - \kappa\left(x_i - \underline{x}\right) z\right] di + o\left(z\right) \\
& = \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}}\int_0^{\underline{i}}\mathbb{E}\left[g_{i,1}(\cdot)\right]\frac{dlog\left[\mathbb{E}\left(u_1'(\cdot)\right)\right]}{dlog\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]} di + \frac{1}{\underline{x}}\beta H v\left(\underline{x}\right) z\int_{\underline{i}}^1 g_{i,1}\left[\mathbb{E}(\cdot)\right] di + o\left(z\right),
\end{aligned}$$

which shows Equation (82). The first-order revenue effect (Equation (83))

$$\begin{aligned}
\mathcal{RE}_{\underline{x}} &= \frac{1}{\underline{x}}\beta H \int_0^{\underline{i}} T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1+\zeta_i^{a,r}\right)\zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} di \\
&+ \frac{1}{\underline{x}}\beta H \int_i T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1+\varepsilon_i^{r,a}\right)\frac{\partial a_i}{\partial I_2}\frac{\partial I_2}{\partial \underline{x}}\frac{\underline{x}}{a_i} di + o\left(z\right) \\
&= \frac{1}{\underline{x}}\beta H \int_0^{\underline{i}} T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1+\zeta_i^{a,r}\right)\zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} di \\
&+ \frac{1}{\underline{x}}\beta H \int_i \frac{T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right]}{1-T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right]} a_i \mathbb{E}\left(r_i^p z\right)\left(1+\varepsilon_i^{r,a}\right)\eta_i^{a,I_2} v\left(\underline{x}\right) z di + o\left(z\right)
\end{aligned}$$

collects the effects on the capital income of households below $\underline{i}$ and income effects for all households.