UniCredit Foundation

"Redistribution of Return Inequality”

Karl Schulz

8th Econ Job Market Best Paper Award

Working Paper Series

n. 166 ■ December 2021

Statement of Purpose

The Working Paper series of the UniCredit Foundation is designed to disseminate and to provide a platform for discussion of either work of the UniCredit economists and researchers or outside contributors (such as the UniCredit Foundation scholars and fellows) on topics which are of special interest to the UniCredit Group. To ensure the high quality of their content, the contributions are subjected to an international refereeing process conducted by the Scientific Committee members of the Foundation.

The opinions are strictly those of the authors and do in no way commit the Foundation and UniCredit.

These Working Papers often represent preliminary work. Citation and use of such a paper should take account of its provisional character.

The Working Papers are also available on our website (http://www.unicreditfoundation.org)

Redistribution of Return Inequality*

Karl Schulz†

University of Mannheim

October 14, 2021

Click here for the most recent version.

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 correlate with ability types and wealth, giving rise to type and scale dependence. Whereas an increase in type dependence ceteris paribus raises optimal capital taxes, more scale dependence provides a rationale for lower taxes, making the policy implications of return inequality non-trivial. The intuition is that, aside from amplifying capital inequality, scale dependence generates an inequality multiplier effect between wealth and its pre-tax return rate. This effect scales up standard elasticity measures that determine the responsiveness of capital to taxes. In a financial market microfoundation, in addition to type and scale dependence, I identify general equilibrium effects that call for more redistribution relative to the partial equilibrium. Finally, I provide macro and micro estimates of the novel sufficient statistics and demonstrate their quantitative importance for capital taxation.

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 am grateful to Eckhard Janeba and Nicolas Werquin for their invaluable guidance and comments throughout this project. I also thank Klaus Adam, Antoine Camous, Ashley Craig, Eduardo Dávila, Sebastian Findeisen, Hans Peter Grüner, 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, 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 U.S., for example, the wealth share of the top 0.1% has tripled over the past forty years (Saez and Zucman (2016)). Persistent heterogeneity in the idiosyncratic returns to wealth has 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 or investment talent. In addition to type dependence, a recent wave of empirical papers documents the prevalence of "scale dependence," referring to a positive correlation between wealth and its return.[1] Scale dependence may have 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 well-known result in public finance is that exogenous inequality in return rates justifies the taxation of capital (see, e.g., Atkinson and Stiglitz (1976) and Saez (2002)). However, little is known about the policy implications of the sources of return inequality. How should capital taxation account for the presence of scale and type dependence? How should redistribution respond to a rise in inequality driven by return rates? Which sources of return inequality should a government address, and which not? Can the government alter the distribution of pre-tax return rates? To answer these questions, I study capital taxation when the inequality in return rates comes from type and scale dependence. As a leading example, I microfound scale dependence on a financial market with portfolio choice and information acquisition following the argument by Piketty (2014), as originated by Arrow (1987).

According to conventional wisdom, one might expect that when the rich experience a rise in their return rates relative to the poor, this additional source of inequality provides a rationale for higher capital taxes (see, for example, Piketty (2014)). However, this paper shows that, besides the inequality level, the source of return inequality is crucial for capital taxation: More type dependence calls for higher capital taxes, whereas scale dependence reduces the optimal capital tax.

The intuition is as follows. The optimal capital tax with return inequality features

[1] For recent empirical evidence, see Bach, Calvet, and Sodini (2020) and Fagereng, Guiso, Malacrino, and Pistaferri (2020). Moreover, it has been shown that one needs to add scale dependence to standard random growth models to account for the cross-sectional dynamics in inequality (Gabaix, Lasry, Lions, and Moll (2016)).

a classic trade-off between equity and efficiency. Firstly, it is inversely related to the elasticity of capital with respect to capital taxes, capturing the efficiency costs of raising taxes. Secondly, the optimal tax rises in the observed capital income inequality, which accounts for equity concerns. This trade-off is present irrespective of the source of return inequality (type and scale dependence).

However, these two measures that determine the optimal capital tax structurally depend on the source of return inequality. Type dependence amplifies the observed capital inequality, which calls for a higher capital tax. Scale dependence also increases the observed amount of capital inequality. At the same time, unlike type dependence, scale dependence raises the elasticity of capital due to a novel efficiency channel.

Under scale dependence, there is a two-way interplay between taxes and pre-tax returns: On the one hand, return rates and their distribution across households shape the wealth distribution, which serves as a critical primitive for designing a tax system. This standard channel is present in taxation models with exogenous return inequality (type dependence only). On the other hand, if the capital elasticity is non-zero, taxes affect the incentives to save. However, in the presence of scale dependence, higher savings boost pre-tax return rates, yielding a convex relationship between capital income and savings. Thus, scale dependence makes pre-tax return rates endogenous to the tax system (novel channel). I demonstrate that this convexity under scale dependence generates an inequality multiplier effect that augments the standard income and substitution effects from tax reforms.[2]

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 savings and its pre-tax return endogenously correlate, the latter also rises. In the leading example, the individual acquires more information, e.g., via financial advisory or financial education, and adjusts the financial portfolio. Now, she earns more on every dollar she invests in future consumption. In other words, saving money pays off to a greater extent. In response, the individual saves more, which, in turn, increases her pre-tax return and so on. The own-return elasticity measures this inequality multiplier effect as it describes the responsiveness of pre-tax returns.

I show that the increase in the elasticity either offsets or dominates the rise in

[2] Scheuer and Werning (2017) demonstrate a similar result for superstar compensation schemes and labor income taxation.

inequality. Perhaps surprisingly, a rise in scale dependence is, thus, either neutral or calls for a lower capital tax. In contrast, more type dependence leads to a higher optimal capital tax since it only raises the observed capital income inequality while leaving the capital elasticity unaltered, as described. Therefore, type and scale dependence have opposite policy implications. It does not only matter for tax policy whether and to which extent return rates are heterogeneous across the population, but the underlying source of return inequality is crucial for understanding how a government should respond to rising inequality. 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).

The endogeneity of pre-tax return rates implies a Le Chatelier principle for capital (see Samuelson (1947)). Under endogenous pre-tax return rates, capital responds more elastically for flexible than for fixed return rates. As a result, an econometrician underestimates capital elasticities if she does not account for the adjustment in pre-tax return rates, for instance, by using data from a short time window in which return rates do not adjust. Even estimates from long-run data may be biased. For example, a bias materializes if one estimates elasticities from the behavior of households in the wealth distribution that do not (or only to a limited extent) feature endogenous returns. This issue is, for example, characteristic of households from the bottom of the wealth distribution who mostly hold cash and are unable to participate in the stock market due to financial constraints.

Therefore, I estimate the amount of scale dependence directly. I first provide reduced-form macro evidence by tracking the relationship between the realized return rate of the rich and their wealth in the Survey of Consumer Finances (SCF). Then, I estimate the own-return elasticity from panel data on the returns of U.S. private foundations. This idea is similar to Piketty (2014), who descriptively documents the amount of scale dependence using return data from U.S. 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. For comparison, I also retrieve an estimate from the study by Fagereng et al. (2020) from Norway.

This procedure yields a range of estimates of the lifetime own-return elasticity between 0.1 and 0.9. A medium value of 0.5, for instance, means that doubling the amount of savings raises a household's rate of return accumulated over a lifetime by

50%. The decline in the optimal capital tax that this magnitude of scale dependence triggers is sizable. For an own-return elasticity of 0.5, the presence of scale dependence lowers the revenue-maximizing capital gains tax by more than 25%.

Aside from illustrating this inequality multiplier effect on the capital elasticity, I derive a novel, parsimonious representation of the measure of inequality that enters the formula for the optimal capital tax. The representation relates well-known inequality measures, such as the Gini coefficient of the capital income distribution, to the distribution's empirically observable shape parameter frequently considered in the inequality literature. I demonstrate how to adjust the inequality measure for scale and type dependence and calculate optimal capital taxation conditional on primitives. I also study the policy implications of the relative amount of scale and type dependence. A boost in inequality can be completely neutral for optimal taxation. For instance, holding the capital income tax fixed at 50%, doubling the own-return elasticity cancels out a surge in type dependence of 17%.

These observations hold under the partial equilibrium assumption of a small open economy and in general equilibrium. However, besides altering individual choices, in general equilibrium, tax reforms also affect aggregate variables that feed back into individual return rates. In the financial market example, the equilibrium stock price aggregates information and risk-taking that depend on aggregate wealth. Thus, an individual's pre-tax return on investment is not only a function of her own savings but also of others'. Then, aside from an inequality multiplier effect, tax reforms also induce inter-household effects. The reasoning is as follows. A tax reform changes an individual'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 individual. I measure these general equilibrium effects in terms of novel cross-return elasticities. In the financial market microfoundation, I identify general equilibrium price effects that call for higher taxes in general than in partial equilibrium. Intuitively, these price effects resemble trickle-up externalities, as in a situation of rent-seeking where the rich take away return rates from the poor. The government taxes these extra rents away (see Rothschild and Scheuer (2016)).

In the foundation data, I find statistically significant but economically small cross-return elasticities. This finding suggests no or negligible general equilibrium forces. The point estimates of cross-return elasticities support some features of the financial market. For instance, there are negative effects from the top of the wealth distribution, indicating the presence of small general equilibrium effects.

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 agents and study the equity and efficiency implications of capital taxation. Gerritsen, Jacobs, Rusu, and Spiritus (2020) analyze capital taxation under type and scale dependence. They show that both sources of return inequality give rise to optimal positive capital taxation and investigate the underlying mechanisms. While my model nests their established (non-)zero-capital-taxation results, I investigate how different sources of return rate inequality shape the equity-efficiency trade-off and demonstrate the opposing effects of scale and type dependence on capital taxation.

Moreover, I introduce scale and type dependence into two well-known taxation 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 and type dependence should carry over to a fully optimal mechanism as considered in the new dynamic public finance literature (for instance, Golosov, Kocherlakota, and Tsyvinski (2003)). Moreover, I abstract from the debate on the gains from selecting different tax policy instruments, e.g., wealth vs. capital income taxation (see Guvenen et al. (2019)) or excess return taxation (for instance, Boadway and Spiritus (2021)).

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, in microfounding return inequality on a financial market, 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)). To the best of my knowledge, this is the first 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, 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)). However, the empirical evidence suggests that scale dependence is widely disseminated throughout the wealth distribution, whereas superstar effects primarily manifest at the top of the income distribution.

Fourthly, in my empirical analysis of the SCF and a large panel of U.S. 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. Finally, by providing estimates of own- and cross-return elasticities, I also add to the empirical literature on estimating capital 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). I also describe the microfoundation and well-known extensions to the conceptual framework. In Section 3, I describe the empirical implications of my results and estimate own- and cross-return elasticities with data from the SCF and U.S. private foundations. In Section 4, I provide quantitative illustrations of the effects of type and scale dependence on the equity-efficiency trade-off and the optimal capital taxation. Section 5 concludes. I relegate all proofs, model extensions, and the microfoundation to the Appendix.

2 The Model

2.1 A Conceptual Framework

This section describes a simple two-period life-cycle framework to think about capital taxation under the presence of type and scale dependence. Suppose there is a unit measure of households $i \in [0,1]$ that differ in their labor earnings ability w_i. Thus, under standard monotonicity conditions, one can interpret i as a household's position in the income distribution. Aside from working l_i hours (in period 1), household i saves a_i (for period 2) to maximize lifetime utility $u_i(c_1, c_2, l)$. Households use their first-period (after-tax) labor income for consumption and savings and consume their final (after-tax) wealth in the second period. Their pre-tax return rates on savings, $r(a_i, i) \equiv r_i(a_i)$, may differ due to *type dependence* and *scale dependence*.[3] Type dependence refers to an exogenous heterogeneity in return rates ($\frac{\partial r_i(a_i)}{\partial i} > 0$): That is, some types can generate higher return rates than others, for example, because of an inherent investment talent or entrepreneurial skill. Scale dependence refers to a positive relationship between wealth and its return ($\frac{\partial r_i(a_i)}{\partial a_i} > 0$).[4] Observe that the presence of scale dependence does not rule out type dependence and vice versa. Just as in reality, both type and scale dependence may co-occur in this setting. In contrast, I refer to *type dependence only* as a setting where all the return inequality is exogenously given ($\frac{\partial r_i(a_i)}{\partial i} > 0$ and $\frac{\partial r_i(a_i)}{\partial a_i} = 0$). Let there be a linear tax rate τ_K on capital gains $a_{R,i} \equiv a_i r_i(a_i)$ and a lump-sum transfer T.[5] Suppose that utility is quasilinear in the consumption of final wealth. Utility maximization yields each household's Marshallian savings supply function $a_i = \overline{a}_i(\tau_K, r_i(a_i); i)$ and an indirect (present-value) utility $U_i(\tau_K, T)$.[6] Define the elasticity of savings with respect to the capital tax rate as

[3] In this section, I focus on persistent return inequality and disregard the role of luck. In the Appendix, I deal with uncertain return rates.

[4] Later, I microfound the notion of scale dependence on a financial market with portfolio choice and financial knowledge acquisition. Wealthy households acquire more financial information and, thus, obtain higher rates of return on their financial investments than poorer households. This microfoundation 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.

[5] Following the capital taxation literature, I interchangeably use the terms capital income and capital gains. Similarly, I do not differentiate between realized and unrealized returns, which is in practice an important distinction. In the Appendix, I extend the exposition to the nonlinear capital income taxation and wealth taxes.

[6] This representation can also be interpreted as the static equivalent of the steady-state utility in a fully dynamic setting (see Saez and Stantcheva (2018)).

$\zeta_i^{a,(1-\tau_K)} \equiv \frac{dlog(a_i)}{dlog(1-\tau_K)}$ 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, households' 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 $\tilde{\zeta}_i^{a,r} \equiv \frac{dlog(a_i)}{dlog(r_i)}$ measure the responsiveness of savings to the rate of return. The novelty of this paper is to explore the differential equity and efficiency effects of type and 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 $\tilde{\zeta}_i^{a,(1-\tau_K)}$, $\tilde{\zeta}_i^{a,r}$, and $\varepsilon_i^{r,a}$ be constant in this section.[7]

A utilitarian social planner maximizes aggregate welfare at a given budget by optimally choosing the capital tax:

$$\max_{\{\tau_K,T\}} \int_i \mathit{\Gamma}_i U_i\left(\tau_K, T\right) di \quad subject \;\; to \quad \int_i \tau_K a_{R,i} di \geq T + \overline{E}, \tag{1}$$

where $\{\mathit{\Gamma}_i\}_{i\in[0,1]}$ are the household's (weakly decreasing) Pareto weights and $\int_i \mathit{\Gamma}_i di = 1$. In the following, I use this basic framework to study capital taxation under type and scale dependence, holding all the other primitives of the economy fixed (such as the savings elasticities at a given rate of return). I establish three novel findings that I summarize in Proposition 1.

Proposition 1. *Consider the optimal capital gains tax under return rate heterogeneity.*

(a) *When expressed in terms of sufficient statistics, the formula for the optimal capital tax is the same irrespective of the sources of return inequality.*

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

(c) *The optimal capital tax with scale dependence is either the same or lower than without scale dependence (type dependence only). By contrast, the presence of type dependence raises the optimal capital tax.*

Proof. Appendix A. □

Part (*a*). The government's problem yields a Ramsey formula for the optimal capital tax (e.g., Diamond (1975))

$$\frac{\tau_K}{1-\tau_K} = \frac{1}{\zeta^{a_R,(1-\tau_K)}} \mathbb{E}\left[\frac{\left(1-\mathit{\Gamma}_i\right) a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\right], \tag{2}$$

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

where $\overline{\zeta}^{a_R,(1-\tau_K)} \equiv \mathbb{E}\left[\frac{a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\zeta_i^{a_R,(1-\tau_K)}\right]$ and $\mathbb{E}\left[\frac{(1-\Gamma_i)a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\right]$ are the average elasticity of capital income and the observed capital income inequality, respectively. Irrespective of how returns form, the optimal capital income tax is decreasing in the average elasticity of capital income which quantifies the efficiency costs of raising taxes. At the same time, accounting for the society's equity concerns, the optimal tax rises in a measure of the observed capital income inequality, e.g., the Gini coefficient of the capital income distribution (see below). Knowing these two sufficient statistics is enough to characterize the optimal capital tax, which gives part (a) of Proposition 1. However, they depend on both type and scale dependence. Therefore, I now move to a structural approach and analyze how the sources of return inequality affect the two sufficient statistics.

Part *(b)*. I begin with the elasticity of capital. Without scale dependence (conditional on pre-tax return rates), capital income is linear in savings $a'_{R,i}\left(a_i\right)|_{\{r_i\}_{i\in[0,1]}} = r_i$ and $a''_{R,i}\left(a_i\right)|_{\{r_i\}_{i\in[0,1]}} = 0$. With scale dependence, the rate of return is endogenous. This makes capital gains convex in savings $a'_{R,i}\left(a_i\right) = r_i\left(a_i\right)$ and $a''_{R,i}\left(a_i\right) = r'_i\left(a_i\right) > 0$, where I assume that households take their equilibrium pre-tax return rates as given. To gain some intuition, consider an individual i. In a setting with type dependence only, the individual is endowed with an investment skill, allowing 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 $i' > i$, she needs to increase her savings substantially. Under scale dependence, the individual has the same return rate r_i in equilibrium as 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, higher savings allow her to raise the 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 for a given return $\overline{\zeta}^{a_R,(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}} = \tilde{\zeta}_i^{a,(1-\tau_K)}$. 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

$$\overline{\zeta}^{a_R,(1-\tau_K)} = \Phi_i\, \overline{\zeta}^{a_R,(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}} \tag{3}$$

with $\varPhi_i \equiv \frac{1+\varepsilon_i^{r,a}}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}} = \left(1+\varepsilon_i^{r,a}\right)\sum_{n=0}^{\infty}\left(\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}\right)^n > 1$ measuring an inequality multiplier effect. The size of the adjustment is proportional to the inequality multiplier effect $\varPhi_i$. 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 pre-tax 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. $\varPhi_i$ 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. The result resembles the effect of superstar compensation schemes in the context labor income taxation (see Scheuer and Werning (2017)). Thus, scale dependence is in its implications for redistribution similar to a superstar phenomenon.[8]

Part (*c*). How does scale dependence affect optimal capital taxes? On the one hand, as described, scale dependence raises the observed capital gains elasticity, which reduces optimal taxes. On the other hand, the presence of scale dependence has the potential to amplify capital income inequality significantly because an initial level of wealth dispersion generates more return inequality which in turn raises wealth inequality. Recall that the optimal capital income tax is increasing in the observed capital income inequality. Through this channel, one would expect higher taxes aimed at reducing 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 $I\left(\tau_K\right) \equiv \mathbb{E}\left[\frac{(1-\varGamma_i)a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\right]$. With constant elasticities, however, $I'\left(\tau_K\right) = 0$. Therefore, compared to an economy with type dependence that is observationally equivalent in terms of inequality $\left(I\left(\tau_K\right) = I\left(\tilde{\tau}_K\right)\right)$, taxes are lower in the economy with scale dependence because the capital income elasticities are

[8] One can also interpret this finding as a Le Chatelier principle for capital (see Samuelson (1947)): In the long run, when pre-tax return rates adjust, capital responds more elastic than in the short run for fixed return rates.

Own-Return	Compensated Elasticity		
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 ($\tilde{\zeta}_i^{a,r} = 0.5$ and $\Gamma_i = 0$).

higher[9]

$$\frac{\tau_K}{1-\tau_K} = \frac{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}}{1+\varepsilon_i^{r,a}} \frac{\tilde{\tau}_K}{1-\tilde{\tau}_K}. \tag{4}$$

To demonstrate the quantitative importance of endogenous pre-tax 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 capital taxes for different values of $\tilde{\zeta}_i^{a,(1-\tau_K)}$ and $\varepsilon_i^{r,a}$. As usual, the larger the savings elasticity, the lower the optimal capital tax. A 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 reduces the revenue-maximizing tax rate by more than 25% (17 percentage points). In the empirical section, I identify a range of estimates of the lifetime own-return elasticity between 0.1 and 0.9. For $\tilde{\zeta}_i^{a,(1-\tau_K)} = 0.5$, this results in a reduction of the optimal capital gains tax between 6 and 40%.

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, instead of considering two in terms of inequality observationally equivalent economies, I compare an economy with scale dependence to one without scale dependence, holding all primitives other than scale dependence fixed. Aside from the

[9] In the dynastic economy of Section C (Mirrleesian economy of Section G), I show that a similar logic applies to a linear wealth tax (nonlinear capital income tax).

savings elasticities at given return rates, these primitives include the wage distribution and the exogenous part of the return rate distribution that measures type dependence. One can approximate the measure of inequality as

$$I(\tau_K) \approx \frac{1}{1-\bar{i}} \zeta_i^{a_R,(1-i)} \mathbb{COV}(\mathit{\Gamma}_i, i),$$

where $\bar{i}$ denotes the household who earns the average capital income and

$$\zeta_i^{a_R,(1-i)} \equiv \frac{dlog(a_i r_i(a_i))}{dlog(1-i)} = \zeta_i^{a,(1-i)} + \zeta_i^{r,(1-i)} = \mathit{\Phi}_i \tilde{\zeta}_i^{a,(1-i)} + \tilde{\zeta}_i^{r,(1-i)}$$

defines the elasticity of capital income with respect to household rank in the income distribution. As I argue in Section 4, for Pareto distributed capital incomes (see, for instance, Gabaix (2009)), the latter elasticity equals the inverse of the shape parameter. Thus, this characterization provides a novel connection between the literature on optimal capital taxation and the measurement of inequality.[10]

In the context of nonlinear capital taxation (Section G), I obtain an exact version of this approximation by expressing the hazard ratio of the capital income distribution in terms of type and scale dependence. In Section B, I quantitatively analyze the performance of the approximation. The approach readily extends to the presence of income effects. Moreover, it allows me to remain completely agnostic about the underlying processes of return formation. Also, notice the similarity to the linearization of, e.g., steady-state equilibria in macroeconomic models.

The elasticity $\zeta_i^{a_R,(1-i)}$ consists of two terms. The first term $\mathit{\Phi}_i \tilde{\zeta}_i^{a,(1-i)}$ captures the described effect that, under scale dependence, any initial level of wealth inequality translates into a more pronounced degree of capital income inequality relative to a setting without scale dependence. The second term $\tilde{\zeta}_i^{r,(1-i)} \equiv \frac{\partial log(r_i)}{\partial log(1-i)}$ is the reduced form relationship between households' rank and their return rates (conditional on an amount of wealth). Therefore, it directly measures the amount of type dependence in the economy.

Thus, in the absence of type dependence ($\tilde{\zeta}_i^{r,(1-i)} = 0$), the introduction of scale dependence is completely neutral. The rise in inequality cancels the increase in the average capital income elasticity. With type dependence ($\tilde{\zeta}_i^{r,(1-i)} < 0$), scale dependence reduces optimal capital taxes. Altogether, despite its potential to boost wealth

[10] See Simula and Trannoy (2020) for a related attempt in the context of optimal nonlinear labor taxation.

inequality, scale dependence either reduces the optimal capital tax or is entirely neutral (Proposition 1 (c)).[11] One can read this result as a possible justification for why capital taxes (e.g., in the U.S.) have not gone up, although capital income inequality has mounted (see Section 3.2). If this rise in inequality came from scale dependence, one should not tax more. However, if it was driven by type dependence, capital taxation should be higher since type dependence raises capital income inequality while leaving the capital elasticity unaffected. In Section 4, I quantitatively illustrate this insight further.

Atkinson-Stiglitz theorem. Altogether, the optimal capital tax can be expressed in terms of primitives

$$\frac{\tau_K}{1-\tau_K} \approx \frac{1}{1-\bar{i}} \frac{\tilde{\zeta}_i^{a,(1-i)} + \tilde{\zeta}_i^{r,(1-i)}/\Phi_i}{\tilde{\zeta}_i^{a,(1-\tau_K)}} \mathbb{COV}\left(\Gamma_i, i\right). \tag{5}$$

This novel and parsimonious representation provides four conditions under which the zero-capital-taxation result holds/does not hold. Firstly, the optimal capital tax is zero when the capital elasticity diverges $\tilde{\zeta}_i^{a,(1-\tau_K)} \to \infty$. Secondly, one obtains a zero-capital-taxation result when each household's relative rank in the income distribution i is unrelated to the social marginal welfare weight Γ_i such that $\mathbb{COV}\left(\Gamma_i, i\right) = 0$.

Thirdly, there is an optimal zero capital tax in the absence of any initial inequality, $\tilde{\zeta}_i^{r,(1-i)} = 0$ and $\tilde{\zeta}_i^{a,(1-i)} = 0$. $\tilde{\zeta}_i^{a,(1-i)}$ describes the amount of reduced-form wealth inequality measured by the relationship between a household's wealth and the rank in the income distribution. This inequality measure depends on the presence of type dependence and the availability of a nonlinear labor income tax. Without a labor income tax, $\tilde{\zeta}_i^{a,(1-i)} = \tilde{\zeta}_i^{a,lw} \tilde{\zeta}_i^{lw,(1-i)} + \tilde{\zeta}_i^{a,r} \tilde{\zeta}_i^{r,(1-i)}$, such that, absent of wage inequality ($w_i = w$, $\forall i$) and type dependence ($r_i = r$, $\forall i$), the optimal capital tax is zero. In this situation, scale dependence does not play a role since there is no underlying wealth inequality translating into endogenous return inequality. With a nonlinear labor income tax, $\tilde{\zeta}_i^{a,(1-i)} = \left(1 - T_l'\left(l_i w_i\right)\right) \tilde{\zeta}_i^{a,lw-T_l(lw)} \tilde{\zeta}_i^{lw,(1-i)} + \tilde{\zeta}_i^{a,r} \tilde{\zeta}_i^{r,(1-i)}$. Then, absent of any type dependence ($\tilde{\zeta}_i^{r,(1-i)} = 0$), the household heterogeneity is, even in the presence of scale dependence, one-dimensional (in terms of wage inequality w_i), and the capital tax is redundant as an instrument (see Atkinson and Stiglitz (1976)).[12]

[11] In Section G, I derive a similar neutrality result for an optimally set nonlinear capital gains tax in a life-cycle economy with nonlinear labor income taxes.

[12] However, the timing of the labor income tax plays a role. If labor and capital incomes are taxed in the same period, there is no role for capital gains taxation (e.g., Gahvari and Micheletto (2016)). When the taxes are levied in different periods and the government cannot freely borrow and save,

Whenever there is a minimal amount of type dependence, the government wishes to levy a non-zero capital tax even if a nonlinear labor income tax is available.[13] In this case, the formula shows that the optimal capital tax rate depends on the relative magnitude of type and scale dependence $\tilde{\zeta}_i^{r,(1-i)}/\Phi_i$. This leads to the fourth condition. When the own-return elasticity converges to the inverse savings elasticity $\varepsilon_i^{r,a} \to \frac{1}{\zeta_i^{a,r}}$, the inequality multiplier effect diverges $\Phi_i \to \infty$. This observation is an entirely novel zero-capital-taxation result. In this case, both the capital elasticity and the observed capital income inequality go to infinity, but the elasticity diverges faster.

Corollary 1. *Under scale dependence, a rise in capital taxation compresses the distribution of pre-tax returns. However, this compression effect comes along with the cost of lowering mean pre-tax returns.*

Proof. Appendix A.4. □

Interestingly, under scale dependence, 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 pre-tax returns is associated with the cost of diminishing mean pre-tax returns

$$d\mathbb{E}(r_i) = -\mathbb{E}(r_i)\,\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)}\frac{d\tau_K}{1-\tau_K} < 0.$$

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. But, on the other hand, there are novel efficiency costs from lowering the level of pre-tax returns.

Gerritsen et al. (2020) find a positive capital tax.

[13]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)).

For fully type-dependent rates of return, only the distribution of after-tax returns but not the pre-tax return distribution responds 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. However, 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 3.2, I use mean responses of the top 1% wealth group to identify scale dependence.

Proposition 2. *In the financial market microfoundation, general equilibrium price effects provide a force for a higher capital gains tax in general than in partial equilibrium.*

Proof. Appendix A.5. □

In Appendix E, I microfound the notion of scale dependence (for a short description, see Section 2.2). On a financial market, households optimally choose their portfolio and the amount of information they wish to acquire. Wealthier households invest more and, thus, have a higher incentive to gain 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)).[14]

I show that for linear costs of information acquisition and when everyone acquires knowledge, a change in the savings by a household i' leads to the same percentage change in the return rate of any other household i $\gamma_{i,i'}^{r,a} = \frac{1}{r_i}\delta_{i'}^{r,a}$. Moreover, $\delta_{i'}^{r,a}$ is decreasing $a_{i'}$. The semi-elasticity is positive for small values of $a_{i'}$ and negative for

[14]Instead of studying a financial market, one could also consider a standard production function that, for instance, exhibits decreasing returns to scale. Then, a household's return function depends on aggregate capital and, thus, on others' capital supply. In any case, the cross-return elasticity describes this dependence.

large ones. These general equilibrium effects are similar to 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. When the substitution effect dominates the income effect, a tax cut on the rich's capital 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, aggregate information also grows 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 goes down, leading to lower return rates for them ($\delta_{i'}^{r,a} < 0$). Thus, the tax cut on the rich increases their return rates but reduces those of the poor.

The formula for the optimal linear capital tax in general equilibrium is given by

$$\frac{\tau_K}{1-\tau_K} = \frac{1}{\overline{\zeta}^{a_R,(1-\tau_K)}} \mathbb{E}\left[\frac{\left(1 - \mathit{\Gamma}_i \left(1 + \gamma_i^{r,(1-\tau_K)}\right)\right) a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\right], \tag{6}$$

where $\gamma_i^{r,(1-\tau_K)} \equiv \int_{i'} \gamma_{i,i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'$ summarizes the general equilibrium effects. Suppose that the cross-return elasticities average out such that $\int_{i'} \gamma_{i',i'}^{r,a} di' = 0$.[15] Then, one can show that the average capital gains elasticity, $\overline{\zeta}^{a_R,(1-\tau_K)}$, declines relative to the partial equilibrium. Moreover, $\gamma_i^{r,(1-\tau_K)} = \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di' < 0$. Both the general equilibrium effects and the adjustment of the capital income elasticity call for a higher capital tax.

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. Accordingly, this politician 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 optimal general equilibrium tax rate is larger than the one in this self-confirming policy equilibrium $\tau_K^{GE} > \tau_K^{PE}$.

Alternatively, one may approximate the measure of capital income inequality in the formula for the optimal capital tax conditional on primitives as in part (c) (see Appendix A.5). For the given specification, the capital income inequality is higher in general than in partial equilibrium, whereas the capital income elasticity is, again, revised downwards. Consequently, in both comparative statics exercises, general equilibrium forces call for more redistribution in general equilibrium. This result is intuitive

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

because, as with trickle-up effects, cutting the capital tax 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 capital income from the poor. It has been shown that 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 statics exercise only compares capital taxes in partial and general equilibrium. To evaluate whether the presence of endogenous return rates in general equilibrium 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.

One can also interpret this result in connection with the integration of financial markets. As markets become internationally more connected, general equilibrium effects vanish ($\gamma^{r,a}_{i,i'} \to 0$). 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 economy, the optimal capital gains tax and, thus, the level of redistribution declines with the international integration of financial markets. Therefore, the decline in the U.S. capital income taxation in the past decades is consistent with a reduction in general equilibrium effects due to financial market integration.

2.2 Microfoundation, Extensions, and Discussion

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

Microfoundation. To begin, I describe the financial market of Appendix 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.[16] I incorporate taxes into this market and demonstrate the functional form of own- and cross-return elasticities in a linear example. Moreover, I include career effects and explicitly add type dependence. Interestingly, the presence of type dependence affects the distribution of own- and cross-return elasticities in the financial market.

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. Nevertheless, to identify general equilibrium effects, if present at all, the sources of scale dependence are relevant to the extent that they may enter differently into the cross-return elasticities, $\gamma_{i,i'}^{r,a}$.

Discussion and extensions. The paper's message is not that taxes should be lower with return inequality than without. Instead, I analyze the differential policy implications of type and scale dependence. Also, lower taxes in the presence of scale dependence do not mean that the government should let wealth and return inequality grow indefinitely. Firstly, there might be an upper bound on households' long-run return rates, naturally limiting the amount of scale dependence and, thus, the upward adjustment in capital income elasticities. Secondly, 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 Section C, I introduce type and scale dependence into the dynamic bequest taxation framework of Piketty and Saez (2013), in which a government taxes the intergenerational transmission of wealth. The environment features income effects, the presence of labor income taxation, uncertain returns, and dynastic considerations. I demonstrate that Proposition 1, Corollary 1, and Proposition 2 carry over to this setting. However, the formulas for optimal wealth taxation now include aggregate wealth and labor income elasticities as well as distributional parameters of labor income and of received and left-over wealth. Moreover, the sufficient statistics are adjusted by another version of the own-return elasticity $\varepsilon_{i,t}^{1+r,a} \equiv \frac{\partial log[1+r_{i,t}(a_{i,t})]}{\partial log(a_{i,t})}$. Similarly, a modification of the cross-return elasticity $\gamma_{i,i',t}^{1+r,a} \equiv \frac{\partial log\left[1+r_{i,t}\left(a_{i,t},\left\{a_{i',t}\right\}_{i'\in[0,1]}\right)\right]}{\partial log\left(a_{i',t}\right)}$ enters the formula for the optimal wealth tax in general equilibrium.

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

In Section G, I consider the nonlinear taxation of capital income in a canonical two-period life-cycle framework (Farhi and Werning (2010)). Using standard perturbation techniques, I derive the nonlinear incidence of capital taxes in partial and general equilibrium and the optimal tax system. Also, in this environment, Proposition 1 continues to hold. Moreover, by expressing the hazard ratio of the capital income distribution in terms of type and scale dependence, I derive, in the context of optimal nonlinear taxation, an exact version of the approximative formula for the optimal linear tax in part (c) of Proposition 1.

As noted by contributors to the literature (e.g., Guvenen et al. (2019)), a capital income and a wealth tax do not coincide with return heterogeneity. In the Guvenen et al. (2019) framework, 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 of 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 in the setting of Guvenen et al. (2019), there would be additional efficiency gains from wealth taxes because the lower tax on high-return individuals increases their pre-tax returns and further expands the tax base. This paper aims to study the effects of scale and type dependence on redistribution where efficiency is one (but not the only) important dimension.

Furthermore, I deal with other policies such as financial education consistent with the leading financial market example (see Section G). Although such policies may be better suited to address return inequality directly, empirical evidence suggests, nonetheless, a residual amount of return inequality 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 (education costs). Similarly, a government would also face 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, such a fund may give rise to other inefficiencies, for example, agency frictions and diversification limits.

Even if these costs declined substantially, it seems unlikely that scale dependence would vanish. For example, in 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 inequality and analyze tax policy for this given residual inequality.

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 instead of type dependence only. 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 that 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, the rich still need to incur costs 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, further lowering the optimal capital tax from an equity perspective.

At the same time, political economy considerations may counteract this force. For example, suppose the political power in a society is endogenous to an individual's wealth. In that case, the amplification of wealth inequality by scale dependence causes may create a rich elite that either directly influences tax policy by running for a political office or indirectly by lobbying. Thus, as Saez and Zucman (2019) proposed, 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, return inequality may provide a rationale for higher capital taxes irrespective of its source.

3 Empirical Analysis

As described, scale dependence affects optimal tax policy by altering the observed capital inequality and standard elasticity measures. Therefore, in this section, I analyze the empirical implications of scale dependence. First, I describe the conceptual issues posed by scale dependence for estimating capital gains and savings elasticities and revisit estimates from the literature (Section 3.1). Next, Section 3.2 provides reduced-form macro evidence of scale dependence using the Survey of Consumer Finances (SCF). Finally, I directly estimate own- and cross-return elasticities using panel data on U.S. foundations (Section 3.3).

3.1 Empirical Implications of Scale Dependence

Conceptual description. First, I describe the implications of scale dependence for estimating capital income and savings elasticities. The standard procedure to estimate the elasticity of capital ($\tilde{\zeta}_i^{a,(1-\tau_K)}$) or capital income ($\tilde{\zeta}_i^{a_R,(1-\tau_K)}$) is to consider individual- or region-specific time-variation in tax rates and study the effects on capital or capital income. Along these lines, consider a tax reform, $d\tau_K$, and abstract from income effects. Then, under scale dependence, the percentage change in the capital holdings of household (or region) i is given by $\frac{da_i}{a_i} = -\tilde{\zeta}_i^{a,(1-\tau_K)} \frac{d\tau_K}{1-\tau_K} + \zeta_i^{a,r} \frac{dr_i}{r_i}$. Similarly, the change in the household's capital income reads as $\frac{da_{R,i}}{a_{R,i}} = -\tilde{\zeta}_i^{a_R,(1-\tau_K)} \frac{d\tau_K}{1-\tau_K} + (1+\zeta_i^{a,r}) \frac{dr_i}{r_i}$.

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 ($dr_i = 0$) suffer from an omitted variable bias when trying to identify the long-run elasticities. Then, the estimation misses the adjustment of returns ($dr_i > 0$), and the error term has a non-zero expectation, conditional on the covariates, violating a critical identifying assumption in empirical studies. If $\zeta_i^{a,r} > 0$, the 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 and a given financial portfolio. In the longer term, she may react to tax reforms, e.g., by hiring other financial advisers, altering the portfolio allocation, and adjusting the pre-tax return rate.

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 hold them for a long time. 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 cover an extended period. This problem also applies to housing, intangible properties, and other assets whose market value only reveals when sold.

Another issue is extrapolating estimates from one group to another in the wealth distribution, even if they are unbiased for the former group. This non-comparability is because portfolios and their flexibility differ significantly across the population. Households from low parts of the wealth distribution mostly hold cash and do not participate in the stock market. Median families have 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 throughout the wealth distribution (see next sections).

Relation to the empirical literature. Now, I summarize two strands of the empirical literature in the light of the described issues. The first strand 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. Unfortunately, the number of studies that explicitly 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 empirical literature on the capital income elasticity, starting from Feldstein, Slemrod, and Yitzhaki (1980), employ microdata and time-series mainly from the U.S. In this literature, the 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 (forthcoming)). The authors distinguish between transitory and permanent responses. Permanent responses seem to be more relevant for long-run tax policy. However, the estimates also need to account for scale dependence in return rates to apply to long-run capital taxation. 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. Brülhart, Gruber, Krapf, and Schmidheiny (2016) analyze Swiss time-series and microdata. Since capital gains on movable assets are untaxed in Switzerland, they cannot directly observe individual return rates. Interestingly, their capital elasticity estimates appear stable over the (upper parts of the) wealth distribution, suggesting uniform own-return elasticities aside from homogeneous intertemporal substitution and tax evasion responses.

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. The estimates do not represent the entire population since the Danish wealth tax only applies to wealthy households in the observation period. To sum up, this literature pays closer attention to unrealized capital gains, which is natural, given its objective to estimate the wealth elasticity. However, the estimates are not readily generalizable to long-run wealth elasticities across the wealth distribution without knowing the amount of scale dependence.

3.2 Macro Evidence

In the following, I propose two approaches to directly estimate the amount of scale dependence that one can use to accommodate the capital elasticity and inequality measures appearing in optimal tax formulas. As I demonstrate in Sections 2 and 4, adjusting these measures for the estimated magnitude of scale dependence is important.

Survey of Consumer Finances. In the spirit of Corollary 1, I estimate scale dependence using the time series of a population group's average return rate and wealth. I extract the household-level asset data from the SCF for 1949-2016 provided by Kuhn, Schularick, and Steins (2020). The representative repeated cross-section contains detailed information on household wealth, portfolio composition, demographic characteristics, and capital income in the U.S. 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

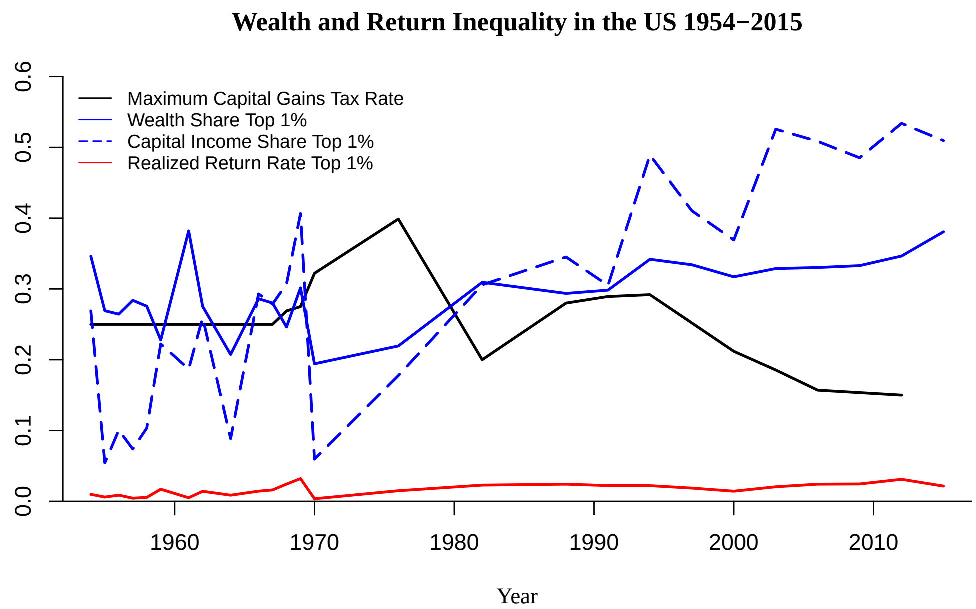

Figure 1: Evolution of Inequality and Capital Taxation in the U.S.

from the wealth concept.

I divide a household's capital income by the invested capital to calculate each household's *realized return rate* in a given year. Note that capital income is reported retrospectively in the sample. Therefore, to avoid reverse causality, I approximate the invested capital in a period by subtracting the past year's capital income from the current year's net wealth. Moreover, I compute a time series of capital income and wealth shares of the top 1% in the wealth distribution. I also add data on capital taxation from the U.S. Department of the Treasury.

Figure 1 displays the evolution of wealth and return inequality in the U.S. over the past decades. The rise in the top 1% capital income share (blue dashed line) is more pronounced than the increase in the respective wealth share (solid blue line). Thus, pre-tax return inequality has grown, consistent with the rising average realized return rate for the top 1% (red line). Simultaneously, the rich experienced a reduction in the capital gains tax rate (black line). Altogether, this suggests that the decline in capital taxation in the U.S. intensified pre-tax return inequality, which is in line with Corollary 1 in Section 2.

Estimation of own-return elasticity. The proposition also suggests using responses of aggregate variables, e.g., mean return rates, to identify scale dependence. In this spirit, I regress the average realized return rate of the top 1% on their log mean wealth, giving a highly significant estimate of 0.01.[17] For an average realized return rate of 1.5%, this estimate translates into a lifetime own-return elasticity of $\widehat{\varepsilon_i^{r,a}} \approx 0.78$ (see below for a more detailed description).

3.3 Micro Evidence

As argued in Section 3.1, using realized return rates to identify scale dependence may be problematic because a substantial share of capital gains is unrealized. Therefore, in the following, I provide evidence from microdata that captures realized and unrealized returns.

Foundation data. I use the publicly available panel data on U.S. 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 U.S. 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, as mentioned, the data set contains both realized and unrealized capital gains, and foundations explicitly report donations and withdrawals.

The main disadvantage of the data set, e.g., compared to the SCF, 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.

[17] With the maximum capital gains tax rate as an instrument, an IV regression yields a similar but at the 5%-level insignificant point estimate.

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: $N = 254\,570$)

As in Fagereng et al. (2020) and Bach et al. (2020), one can directly calculate the investment return of foundation i during a period h, $r_{i,h}$, as the market-value capital income (both 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%.[18] 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 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 a median portfolio size of \$6 978 721. There is a substantial degree of heterogeneity. Foundations differ in their endowment size (wealth inequality) and their investment returns (return inequality). Whereas small foundations (below \$100k) attain an annual return of 2.7%, the top 1% of 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. Accordingly, the amount of scale dependence is likely to be underestimated in the data, and the resulting estimates can be considered

[18] If one leaves out foundation-year observations with return rates below the 2.5th and above the 97.5th percentile, the results will be similar. The uncut sample features a kurtosis (above 100 000) 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.

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

$$log\,(1 + r_{i,h}) = \varepsilon \cdot log\,(a_{i,h}) + f_i + f_h + u_{i,h}, \tag{7}$$

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. There, the own-return elasticity decreases with wealth. To capture these nonlinearities, I estimate an alternative specification

$$log\,(1 + r_{i,h}) = \varepsilon \cdot log\,(a_{i,h}) + \sum_{g'=2}^{7} \varepsilon_{g'} \cdot log\,(a_{i,h}) \cdot D_{g_{i,h},g'} + f_i + f_h + u_{i,h}, \tag{8}$$

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

In Table 3, I report the estimated coefficients of specifications (7) and (8). They reveal a highly significant amount of scale dependence. Doubling a foundation's endowment raises its annual return rate by 0.23 percentage points (annual own-return semi-elasticity). There is no evidence for increasing or decreasing returns to scale. Interestingly, for high foundations sizes, the point estimates of (8) show (slightly non-significant) decreasing returns to scale that would be in line with the financial market example in Section E.2.2. Thus, whereas the specification cannot confirm the parametrization in the financial market, it does neither reject it.

In specifications (9) and (10), I replace log net wealth in (7) by foundation's group affiliation and percentile in the wealth distribution, both based on foundations' beginning-of-year net wealth:

$$log\,(1 + r_{i,h}) = \varepsilon \cdot g_{i,h} + f_i + f_h + u_{i,h}, \tag{9}$$

	Constant Returns to Scale			Incr./Decr. Returns to Scale
	(7)	(9)	(10)	(8)
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; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

and

$$log\left(1 + r_{i,h}\right) = \varepsilon \cdot p_{i,h} + f_i + f_h + u_{i,h}, \tag{10}$$

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 (7) does not change the results qualitatively. Instrumenting foundation wealth in (7) with three-year lagged donations yields the same results. Estimating (7) separately for boom and bust years (1990, 2001, 2008, and 2009) shows 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 (7) into a value for the lifetime own-return elasticity $\varepsilon_i^{r,a}$. Multiply the estimate of (7) 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 lifetime own-return elasticity, consider the compound return rate $R_m = \left(1 + r_{m,h}\right)^H - 1$. Accordingly, one obtains an expression for the lifetime own-return elasticity $\widehat{\varepsilon_i^{r,a}} = \frac{H \cdot \left(1+r_{m,h}\right)^{H-1}}{R_{m,h}} \frac{dlog\left(1+r_{i,h}\right)}{dlog\left(a_{i,h}\right)}$. For $r_{m,h} = 4.9\%$ and $H = 30$, this yields an estimate of $\widehat{\varepsilon_i^{r,a}} \approx 0.1$. Even this conservatively estimated effect leads to a notable adjustment of the optimal capital tax.

Comparison to Fagereng et al. (2020). By comparing the estimate of (10) 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 lifetime own-return elasticity of $\widehat{\varepsilon_i^{r,a}} \approx 0.9$ in their data set,[19] which is higher than the estimate obtained in the microdata but close to the reduced-form macro estimate (Section 3.2).

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 one's own savings but also on those of others. A change in the savings by household i' also affects household i's return $\frac{d(1+r_{i,h})}{1+r_{i,h}} = \varepsilon_{i,h}^{1+r,a} \cdot \frac{da_{i,h}}{a_{i,h}} + \int_{i'} \gamma_{i,i',h}^{1+r,a} \cdot \frac{da_{i',h}}{a_{i',h}} di'$. To bring this formulation closer to the data, consider the discrete counterpart $\frac{d(1+r_{i,h})}{1+r_{i,h}} = \varepsilon_{i,h}^{1+r,a} \cdot \frac{da_{i,h}}{a_{i,h}} + \sum_{i'} \gamma_{i,i',h}^{1+r,a} \cdot \frac{da_{i',h}}{a_{i',h}}$.

In the following, I estimate the magnitude of general equilibrium effects (for each wealth group). To be able to identify cross effects, I impose more structure on these effects. I assume that they are constant over time 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 effects be similar within a wealth group $\delta_{i',h}^{r,a} \approx \delta_{g',h}^{r,a}$ for all $i' \in I_{g'}$ and let $\delta_{g',h}^{r,a}$ be small ($\delta_{g',h}^{r,a} \approx 0$). 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

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

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

For small cross effects ($\delta_{g',h}^{r,a}$) and return rates ($r_{i,h} - \mathbb{E}_g\left(r_{i,h}\right)$) and similar cross-effects in each wealth group ($\delta_{i',h}^{r,a} \approx \delta_{g',h}^{r,a}$), the bias term becomes negligible $u_{i,h} \approx 0$.

[19] 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 lifetime 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 $\widehat{\varepsilon_i^{r,a}} \approx 0.7$.

	$g_{i,h}=1$	$g_{i,h}=2$	$g_{i,h}=3$	$g_{i,h}=4$	$g_{i,h}=5$	$g_{i,h}=6$	$g_{i,h}=7$
$g_{i,h-1}=1$	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: $N = 225\,637$)

Therefore, I specify the econometric model by augmenting (7) with cross effects

$$log\,(1 + r_{i,h}) = \varepsilon \cdot log\,(a_{i,h}) + \sum_{g'=1}^{7} \delta_{g'} \cdot \overline{log\,(a_{g',h})} \cdot g_{i,h} + f_i + f_h + u_{i,h} \tag{11}$$

and, controlling for group-specific effects,

$$log\,(1 + r_{i,h}) = \varepsilon \cdot log\,(a_{i,h}) + \beta \cdot g_{i,h} + \sum_{g'=1}^{7} \delta_{g'} \cdot \overline{log\,(a_{g',h})} \cdot g_{i,h} + f_i + f_h + u_{i,h} \tag{12}$$

where, again, $g_{i,h}$ indicates foundation i's group affiliation and $\overline{log\,(a_{g',h})} \equiv \sum_{i' \in I_{g'}} log\,(a_{i',h})$ 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 (11) and (12). The estimated scale dependence (own-return elasticity) 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 ($\delta^{r,a}_{i,h} \approx 0$) in the comparative statics (Sections 2 and C) and validate the identifying assumption that $u_{i,h} \approx 0$ (for $\delta^{r,a}_{i',h} \approx \delta^{r,a}_{g',h}$).

	Constant Returns to Scale		Constant Returns to Scale	
	(11)		(12)	
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; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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 effects from the top, as in the general equilibrium financial market. Moreover, using the estimated coefficients from specification (12), a simple Wald test does not reject the hypothesis that $\int_{i'} \gamma^{r,a}_{i',i'} di' = 0$ at the 5% level, which is in line with the assumption in the theoretical section.

To account for potential group-specific nonlinearities in cross effects, $\delta_{g',g''}$, I re-estimate equations (11) and (12). Again, the estimated cross effects are economically small. Moreover, the estimates do not reveal noteworthy nonlinearities. 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 close to the reduced-form estimate from the SCF in Section 3.2 (0.8). In all cases, the resulting adjustment of capital elasticities and inequality measures 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 effects. Some of the cross-effects estimates seem to be in line with the specified general equilibrium financial market model. More research is needed to assess how far the estimates from foundations apply to household data and whether general equilibrium effects are present.

4 Quantitative Illustration

Now, I provide a quantitative exploration of the theoretical results in Section 2 for an empirically plausible range of type and scale dependence (see Section 3). First, I illustrate the inequality multiplier effect showing the efficiency effects of scale dependence. Then, I show how to adjust capital income inequality measures for scale and type dependence. Finally, I conduct comparative statics exercises of optimal capital gains taxes with respect to type and scale dependence and demonstrate how their relative magnitude matters for optimal taxation.

4.1 Inequality Multiplier Effect

As I show in Proposition 1 (b), scale dependence gives rise to an inequality multiplier effect that amplifies the capital income elasticity and reduces the optimal capital gains tax. To demonstrate the quantitative importance of this multiplier effect, I consider, as a first exercise, a set of economies that differ in the magnitude of scale dependence measured by the own-return elasticity. From now on, I set the reduced-form savings elasticity with respect to the rate of return to $\tilde{\zeta}_i^{a,r} = 0.5$.

The Left Panel of Figure 2 displays the relationship between unadjusted and adjusted capital income elasticities. The former can be interpreted as the reduced-form elasticity $\overline{\zeta}^{a_R,(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}}$ that describes the short-run responses of capital income, holding return rates fixed. As explained, the adjusted capital elasticity $\overline{\zeta}^{a_R,(1-\tau_K)}$ accounts for the endogeneity of return rates that makes capital more responsive. As one can see from the figure, the upward adjustment for a given unadjusted elasticity is notable even for smaller values of the own-return elasticity. The higher the underlying unadjusted elasticity, the higher this difference. Moreover, for a given adjusted elasticity measure, the unadjusted elasticity substantially declines in the value of scale dependence. Altogether, scale dependence increases the efficiency costs of raising capital taxes.

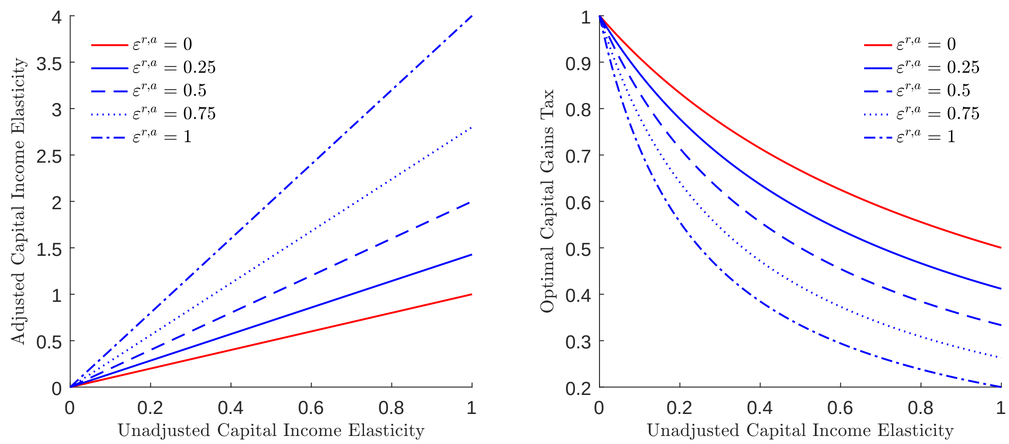

Figure 2: Left Panel: Unadjusted vs Adjusted Capital Income Elasticites; Right Panel: Optimal Rawlsian Capital Gains Taxes

Whereas the measure of capital income inequality, such as the Gini coefficient, is (relatively) easy to observe, the degree to which portfolio returns are scale-dependent and, thus, 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. The Right Panel calculates optimal revenue-maximizing capital gains taxes ($\Gamma_i = 0$) for different values of $\bar{\zeta}^{a_R,(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}}$. For this objective function, the measure of capital income inequality that carries equity concerns plays no role in the optimal capital gains taxation. The red line is the benchmark where are all return inequality is exogenous (type dependence only). One can observe that scale dependence substantially reduces the optimal capital gains tax (blue lines). The higher the unadjusted elasticity and the greater the amount of scale dependence, the larger the adjustment in the optimal capital gains tax. For instance, an own-return elasticity of 0.25 leads to an adjustment in the capital elasticity of more than 42% . For an unadjusted capital income elasticity of 0.25, the resulting scale-dependence induced reduction in the optimal capital gains tax is around 8%.

4.2 Capital Income Inequality

In the previous exercise, I studied the effects of scale dependence on efficiency. However, type and scale dependence also affect a society's equity concerns. In the formula for the optimal capital gains tax, equity concerns are summarized by the inequality measure

$$\mathbb{E}\left[\frac{(1-\mathit{\Gamma}_i)\,a_{R,i}}{\mathbb{E}\,(a_{R,i})}\right] \approx \frac{1}{1-\bar{i}}\zeta_i^{a_R,(1-i)}\mathbb{COV}\,(\mathit{\Gamma}_i, i)\,.$$

This exercise aims to demonstrate how to adjust the measure of capital income inequality by the presence of scale and type dependence.

To illustrate this, I set the Pareto weights to $\mathit{\Gamma}_i = 2\,(1-i)$. This rank-dependent function is known as the Gini social welfare function introduced by Sen (1974). In line with the literature on wealth and income inequality, suppose that capital incomes are Pareto distributed (e.g., Gabaix (2009)). Then, the elasticity of capital income with respect to the household rank is constant $\zeta_i^{a_R,(1-i)} = -1/\lambda_{a_R}$, where λ_{a_R} is the shape parameter of the (steady-state) capital income distribution. Moreover, the inequality measure $I\,(\tau_K)$ is equal to the Gini coefficient. Using the parametrization of welfare weights, $\mathbb{COV}\,(\mathit{\Gamma}_i, i) = -1/6$. Moreover, I set $\bar{i} = 0.8$. This value is consistent with a shape parameter of $\lambda_{a_R} = 1.6$ (see Saez and Stantcheva (2018)). In Appendix B, I analyze the performance of the approximation for different levels of inequality (λ_{a_R}) and alternative social welfare functions ($\mathit{\Gamma}_i$). Moreover, I omit the endogeneity of $\bar{i}$ in this section. In the robustness analysis of Appendix B, I take the endogeneity explicitly into account.

Using this parametrization, I now demonstrate how the inequality measure (Gini coefficient) differs in the short run (unadjusted inequality) and the long run (adjusted inequality). Notice that one can decompose capital income inequality into wealth and return inequality, since $\zeta_i^{a_R,(1-i)} = \zeta_i^{a,(1-i)} + \zeta_i^{r,(1-i)}$. Return inequality $\zeta_i^{r,(1-i)} = \varepsilon_i^{r,a}\zeta_i^{a,(1-i)} + \tilde{\zeta}_i^{r,(1-i)}$ depends on the degree to which it is driven by scale dependence, $\varepsilon_i^{r,a}\zeta_i^{a,(1-i)}$, and a measure of type dependence, $\tilde{\zeta}_i^{r,(1-i)}$, that describes the exogenous differences in return rates. The former depends on the amount of wealth inequality $\zeta_i^{a,(1-i)} = \phi_i\tilde{\zeta}_i^{a,(1-i)} = \frac{\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)}+\tilde{\zeta}_i^{a,r}\tilde{\zeta}_i^{r,(1-i)}}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}}$ that is also a function of type and scale dependence ($\tilde{\zeta}_i^{r,(1-i)}$ and $\varepsilon_i^{r,a}$).

On the left-hand side of Figure 3, I display a situation where all the capital income and wealth inequality is entirely driven by return inequality ($\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = 0$). Without any underlying inequality ($\tilde{\zeta}_i^{r,(1-i)} = 0$), both the adjusted and the unadjusted inequality measures and, thus, the optimal capital tax rates are equal to zero for any amount of scale dependence. A positive amount of type dependence gives rise to capital income inequality. With scale dependence ($\varepsilon^{r,a} > 0$), the adjusted inequality is larger than the unadjusted one. This adjustment is moderate in the absence of any

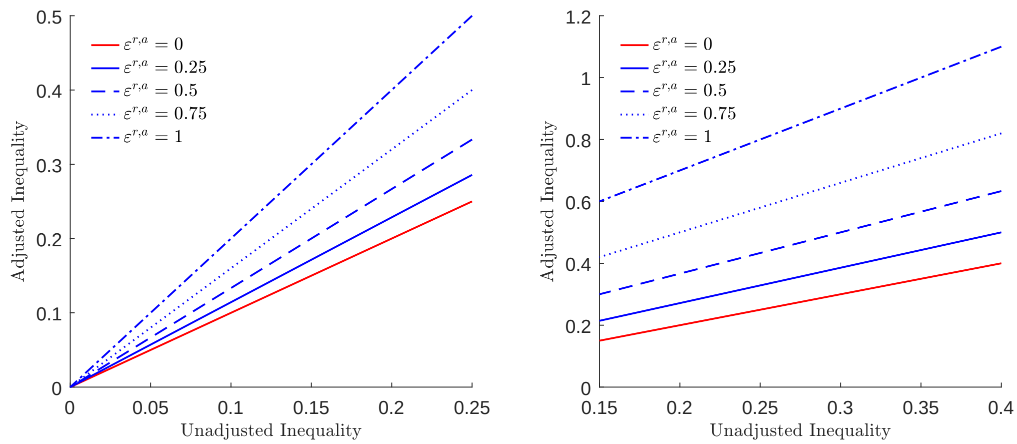

Figure 3: Left Panel: Unadjusted vs Adjusted Inequality ($\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = 0$); Right Panel: Unadjusted vs Adjusted Inequality ($\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = -0.18$)

additional source of inequality ($\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = 0$).

However, the difference between the two inequality measures is more sizable when there is another source of wealth inequality, as shown in the Right Panel, where I set $\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = -0.18$.[20] Therefore, a high long-run Gini coefficient of capital income (adjusted inequality) is not necessarily driven by growing type dependence (unadjusted inequality). Under scale dependence, the rich become richer because they are rich. Their accumulation of wealth makes capital income more unequal in the long run than in the short run. This effect is more substantial when there is more underlying inequality.

To provide an example, without an underlying inequality ($\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = 0$), an own-return elasticity of 0.25 adjusts the Gini coefficient from 0.2 to 0.23. For $\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = -0.18$, the adjustment is more pronounced (from 0.2 to 0.27).

[20] I obtain this value in two steps. First, I calculate the shape parameter of the return rate distribution $\zeta_i^{r,(1-i)}$. In line with Saez and Stantcheva (2018), let $\zeta_i^{a_R,(1-i)} = -1/1.6$. Moreover, I set $\zeta_i^{a,(1-i)} = -1/\lambda_a = -1/3$, which is consistent with the fact that the top 10% wealth share is twice the top 1% wealth share and four times the top 0.1% wealth share (see, for instance, Piketty (2014)). Accordingly, $\zeta_i^{r,(1-i)} = \zeta_i^{a_R,(1-i)} - \zeta_i^{a,(1-i)} \approx -0.3$. Secondly, I assume that, in the baseline scenario, half of the return inequality is driven by type dependence, $\tilde{\zeta}_i^{r,(1-i)} = \frac{1}{2}\zeta_i^{r,(1-i)}$, and the other half is due to scale dependence, $\varepsilon_i^{r,a}\zeta_i^{a,(1-i)} = \frac{1}{2}\zeta_i^{r,(1-i)}$. This assumption implies an own-return elasticity of $\varepsilon_i^{r,a} \approx 0.45$ that is just in the middle of the range identified in Section 3. Moreover, it gives the value of $\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} \approx -0.18$.

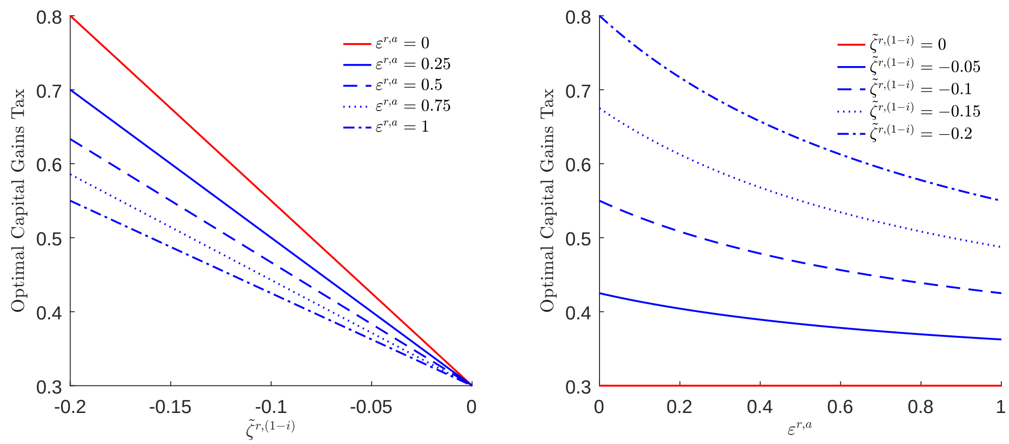

Figure 4: Left Panel: Optimal Capital Income Taxes vs Type Dependence; Right Panel: Optimal Capital Income Taxes vs Scale Dependence

4.3 Optimal Capital Gains Taxation

Altogether, scale dependence raises the elasticity of capital income and amplifies capital income inequality. Although type dependence also contributes to capital income inequality, it does not alter the capital income elasticity. Therefore, the implications for tax policy of rising return inequality are non-trivial.

In this numerical exercise, I explore the role of type and scale dependence for the optimal taxation of capital gains, holding all other primitives fixed. This illustrates Proposition 1 (c). Again, let $\tilde{\zeta}_i^{a,r} = \tilde{\zeta}_i^{a,(1-\tau_K)} = 0.5$, $\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)} = -0.18$, $\bar{i} = 0.8$, and $\Gamma_i = 2\,(1-i)$.

In the Left Panel of Figure 4, I present the optimal capital gains tax as a function of the reduced-form return inequality that measures type dependence, $\tilde{\zeta}_i^{r,(1-i)}$. As argued theoretically, more type dependence (smaller $\tilde{\zeta}_i^{r,(1-i)}$) translates into a greater capital income inequality. Thus, a rise in inequality that is induced by type dependence calls for higher capital taxation. However, this adjustment in the optimal tax is weaker the greater the magnitude of scale dependence.

For example, consider an increase in type dependence from -0.1 to -0.2. Without scale dependence, this leads to a rise in the optimal capital gains tax by 45%. For an own-return elasticity of 0.25, the respective increase is 40%.

The Right Panel of the figure shows the optimal capital tax for different values of scale dependence, measured by the own-return elasticity. A rise in inequality triggered by more scale dependence tends to reduce the optimal capital gains tax rate.

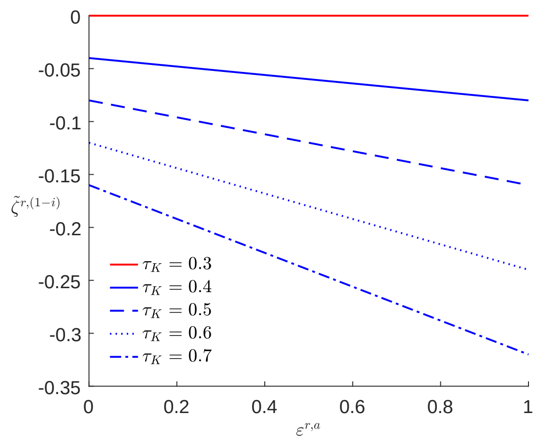

Figure 5: Type vs Scale Dependence

The adjustment due to scale dependence depends on the amount of type dependence. Without type dependence, there is no adjustment (full neutrality of scale dependence, red line). Thus, the larger the underlying type dependence, the stronger is the reduction in optimal capital income taxes due to a scale dependence-induced rise in return inequality. For example, given a type dependence of $\tilde{\zeta}_i^{r,(1-i)} = -0.1$, a rise in $\varepsilon^{r,a}$ from 0.2 to 0.4 lowers the optimal capital gains tax by around 6%. This reduction is more pronounced (more than 8%) when type dependence is, for instance, $\tilde{\zeta}_i^{r,(1-i)} = -0.2$.

To further explore the relative importance of scale and type dependence, I finally compute the isoquants of the optimal tax function in Figure 5. Consider a rise in capital income inequality that is driven by both type and scale dependence. The figure reveals that a rise in inequality does not necessarily alter optimal capital taxation. There are combinations of type and scale dependence for which rising inequality is entirely neutral for tax policy. For example, at an optimal capital tax of 50%, a rise in scale dependence from $\varepsilon^{r,a} = 0.2$ to $\varepsilon^{r,a} = 0.4$ cancels out an increase in type dependence by 17%.

In this section, I quantitatively explored the consequences of scale and type dependence for common sufficient statics, such as the elasticity and inequality of capital income, as well as the optimal capital taxation. Altogether, these measures may substantially differ in the short and long run due to the endogeneity of pre-tax return rates. Moreover, type and scale dependence have opposing effects on optimal taxation. While a higher scale dependence tends to reduce taxes because the adjustment in the efficiency costs dominates the inequality rise, type dependence raises optimal taxes as the efficiency channel is absent. Overall, one cannot infer from an increase in capital income inequality that capital taxes should rise. Depending on the source of this inequality rise, it can have qualitatively opposing policy implications. This difference can be quantitatively considerable.

5 Conclusion

This paper introduces scale and type dependence into the optimal taxation of capital. I show that it does not only matter if and how much return inequality there is, but the source of inequality is essential for tax policy. Both type and scale dependence raise capital inequality. Scale dependence, however, makes capital more elastic to tax reforms, expanding the efficiency costs of capital taxation. I show how to adjust standard sufficient statistics that determine the capital elasticity for scale dependence. These need to account for an inequality multiplier effect between wealth and its pre-tax return.

When scale dependence raises inequality, optimal capital taxes decline because the inequality multiplier effect offsets the increase in the observed level of inequality. Conversely, capital tax rates should increase if the same rise in inequality were driven by type dependence. As a consequence, a government should address changes in capital inequality very differently depending on their source. Therefore, exploring the magnitude of scale dependence relative to type dependence is an important avenue for future research.

References

Aaberge, R. (2000). "Characterizations of lorenz curves and income distributions." *Social Choice and Welfare*, *17*(4), 639–653.

Agersnap, O., and Zidar, O. M. (forthcoming). "The tax elasticity of capital gains and revenue-maximizing rates." *American Economic Review: Insights*.

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." *Unpublished Manuscript, Williams College*.

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.

Boadway, R., and Spiritus, K. (2021). "Optimal taxation of normal and excess returns to risky assets." *Tinbergen Institute Discussion Paper 2021-025/VI*.

Brülhart, M., Gruber, J., Krapf, M., and Schmidheiny, K. (2016). "Taxing wealth: Evidence from switzerland." *National Bureau of Economic Research Working Paper No. 22376*.

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." *Annual Review of Economics*, *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. (2009). "Power laws in economics and finance." *Annu. Rev. Econ.*, *1*(1), 255–294.

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

Gahvari, F., and Micheletto, L. (2016). "Capital income taxation and the atkinson–stiglitz theorem." *Economics Letters*, *147*, 86–89.

Gerritsen, A., Jacobs, B., Rusu, A. V., and Spiritus, K. (2020). "Optimal taxation of capital income with heterogeneous rates of return." *CESifo Working Paper No. 8395*.

Golosov, M., Kocherlakota, N., and Tsyvinski, A. (2003). "Optimal indirect and capital taxation." *The Review of Economic Studies*, *70*(3), 569–587.

Golosov, M., Tsyvinski, A., and Werquin, N. (2014). "A variational approach to the analysis of tax systems." *National Bureau of Economic Research Working Paper No. 20780*.

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 Working Paper No. 26284*.

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.

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.* Harvard University Press Cambridge, MA.

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. (1947). "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.

Sen, A. (1974). "Informational bases of alternative welfare approaches: aggregation and income distribution." *Journal of Public Economics*, *3*(4), 387–403.

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

Simula, L., and Trannoy, A. (2020). "Gini and optimal income taxation by rank." *CESifo Working Paper No. 8141.*

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.

U.S. Department of the Treasury (2016). "Taxes paid on capital gains for returns with positive net capital gains: 1954-2014." `https://home.treasury.gov/policy-issues/tax-policy/office-of-tax-analysis`, accessed: 2021-04-11.

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

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." *Norwegian School of Economics Working Paper.*

Online Appendix for "Redistribution of Return Inequality"

A Proofs of Section 2

A.1 Part (a) of Proposition 1

With and without scale dependence, the government solves $\underset{\tau_K}{max} \int_i \mathit{\Gamma}_i U(\tau_K; i)\, di$ 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 $\mathcal{L} = \int_i \mathit{\Gamma}_i U(\tau_K; i)\, di + \lambda \left[\int_i \tau_K a_{R,i} di - \overline{E}\right]$ with respect to τ_K, the first-order condition reads as

$$\int_i (\mathit{\Gamma}_i/\lambda) \frac{dU(\tau_K; i)}{d\tau_K} di + \int_i a_{R,i} di = \frac{\tau_K}{1-\tau_K} \int_i a_{R,i} \zeta_i^{a_R,(1-\tau_K)} di. \tag{13}$$

With a utility function that is quasilinear in the consumption of final wealth, the first-order effect on household utility is given by $\frac{dU(\tau_K;i)}{d\tau_K} = -a_{R,i}$ and the shadow value of public funds is equal to $\lambda = \int_i \mathit{\Gamma}_i di = 1$. Simplify (13) 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

$$\overline{\zeta}^{a_R,(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}} = \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} \tilde{\zeta}_i^{a_R,(1-\tau_K)} di = \tilde{\zeta}_i^{a_R,(1-\tau_K)} = \tilde{\zeta}_i^{a,(1-\tau_K)} \tag{14}$$

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

$$\begin{aligned}\zeta_i^{a,(1-\tau_K)} &\equiv \frac{dlog(a_i)}{dlog(1-\tau_K)} = \frac{dlog(a_i)}{dlog(1-\tau_K)}|_{\{r_i\}_{i\in[0,1]}} + \frac{dlog(a_i)}{dlog(r_i)} \frac{dlog[r_i(a_i)]}{dlog(a_i)} \frac{dlog(a_i)}{dlog(1-\tau_K)} \\ &= \tilde{\zeta}_i^{a,(1-\tau_K)} + \tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}\zeta_i^{a,(1-\tau_K)} = \phi_i\tilde{\zeta}_i^{a,(1-\tau_K)}\end{aligned}$$

and the capital income elasticity

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

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

$$\overline{\zeta}^{a_R,(1-\tau_K)}=\int_i \frac{a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,(1-\tau_K)}di=\frac{1+\varepsilon_i^{r,a}}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}}\tilde{\zeta}_i^{a,(1-\tau_K)} \tag{15}$$

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

A.3 Part (c) of Proposition 1

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

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

For constant elasticities $\tilde{\zeta}_i^{a,(1-\tau_K)}$, $\tilde{\zeta}_i^{a,r}$, 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 $I'\left(\tau_K\right)$ is equal to zero.

Define $\bar{i}$ such that $a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)=\mathbb{E}\left(a_i r_i\left(a_i\right)\right)$. Then, approximate each household's capital income around the one of household $\bar{i}$, $a_i r_i=a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)-\frac{da_i r_i}{d(1-i)}\frac{1-\bar{i}}{a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)}\frac{i-\bar{i}}{1-\bar{i}}a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)+o\left(i-\bar{i}\right)$, such that $\frac{a_i r_i-a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)}{a_{\bar{i}}r_{\bar{i}}\left(a_{\bar{i}}\right)}=\frac{a_{R,i}-\mathbb{E}\left(a_{R,i}\right)}{\mathbb{E}\left(a_{R,i}\right)}=-\zeta_i^{a_R,(1-i)}\frac{i-\bar{i}}{1-\bar{i}}+o\left(i-\bar{i}\right)$, where $\zeta_i^{a_R,i}\equiv\frac{dlog\left(a_{R,i}\right)}{dlog(1-i)}$. Therefore, for constant elasticities,

$$\begin{aligned}I\left(\tau_K\right)&=\mathbb{E}\left[\left(1-\mathit{\Gamma}_i\right)\frac{a_{R,i}}{\mathbb{E}\left(a_{R,i}\right)}\right]=\mathbb{E}\left[\left(1-\mathit{\Gamma}_i\right)\frac{a_{R,i}-\mathbb{E}\left(a_{R,i}\right)}{\mathbb{E}\left(a_{R,i}\right)}\right]\\ &\approx-\frac{1}{1-\bar{i}}\zeta_i^{a_R,(1-i)}\mathbb{E}\left[\left(1-\mathit{\Gamma}_i\right)\left(i-\bar{i}\right)\right]=\frac{1}{1-\bar{i}}\zeta_i^{a_R,(1-i)}\mathbb{COV}\left(\mathit{\Gamma}_i,i\right).\end{aligned}$$

Moreover, the wealth elasticity with respect to the household rank can be written as

$$\begin{aligned}\zeta_i^{a,(1-i)}=&\frac{dlog\left(a_i\right)}{dlog\left(R_{1,i}\right)}\frac{dlog\left(R_{1,i}\right)}{dlog\left(1-i\right)}|_{\{r_i\}_{i\in[0,1]}}+\frac{dlog\left(a_i\right)}{dlog\left(r_i\right)}\frac{dlog\left(r_i\left(a_i\right)\right)}{dlog\left(1-i\right)}|_{\{r_i\}_{i\in[0,1]}}\\ &+\frac{dlog\left(a_i\right)}{dlog\left(r_i\right)}\frac{dlog\left(r_i\left(a_i\right)\right)}{dlog\left(a_i\right)}\frac{dlog\left(a_i\right)}{dlog\left(1-i\right)}=\phi_i\left(\tilde{\zeta}_i^{a,R_1}\tilde{\zeta}_i^{R_1,(1-i)}+\tilde{\zeta}_i^{a,r}\tilde{\zeta}_i^{r,(1-i)}\right)=\phi_i\tilde{\zeta}_i^{a,(1-i)},\end{aligned}$$

where $R_{1,i}$ is a household's first-period (after-tax) labor income. Hence, the respective capital income elasticity is given by

$$\zeta_i^{a_R,(1-i)} = \frac{dlog\left(a_i r_i\left(a_i\right)\right)}{dlog\left(1-i\right)} = \left(1+\varepsilon_i^{r,a}\right)\zeta_i^{a,(1-i)} + \frac{\partial log\left(r_i\left(a_i\right)\right)}{\partial log\left(1-i\right)}|_{\{r_i\}_{i\in[0,1]}} = \Phi_i \tilde{\zeta}_i^{a,(1-i)} + \tilde{\zeta}_i^{r,(1-i)}$$

and the optimal tax rate reads as

$$\frac{\tau_K}{1-\tau_K} \approx \frac{1}{1-\bar{i}} \frac{\Phi_i \tilde{\zeta}_i^{a,(1-i)} + \tilde{\zeta}_i^{r,(1-i)}}{\Phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}} \mathbb{COV}\left(\Gamma_i, i\right).$$

Note that, for given elasticities and Pareto weights, this formula is not entirely in closed form since the household that earns the mean capital income $\bar{i}$ is an endogenous variable. I omit this endogeneity for simplicity.

A.4 Corollary 1

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

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

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

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

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

A.5 Proposition 2

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

$$\int_i \left(\Gamma_i/\lambda\right) \frac{dU\left(\tau_K; i\right)}{d\tau_K} di + \int_i \left(\Gamma_i/\lambda\right) \frac{dU\left(\tau_K; i\right)}{dr_i} \int_{i'} \frac{dr_i}{da_{i'}} \frac{da_{i'}}{d\tau_K} di' di + \int_i a_{R,i} di = \frac{\tau_K}{1-\tau_K} \int_i a_{R,i} \zeta_i^{a_R,(1-\tau_K)} di, \tag{16}$$

where the second term on the left-hand side of (16) collects cross-effects in each households' return rates. Note that by the quasilinearity of the utility function

$\frac{dU(\tau_K;i)}{dr_i} = (1-\tau_K)\, a_i$. Using the definition of cross-return elasticities, the first-order inter-household effects simplify to

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

leading to the optimal capital gains tax in general equilibrium.

Elasticities in general equilibrium. Observe that, aside from collecting general equilibrium effects, 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

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

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

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

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

$$\zeta_i^{a,(1-\tau_K)} = \left(1 + \tilde{\zeta}_i^{a,r} \phi_i \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} di'\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} < \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}. \tag{17}$$

Notice that the savings elasticity is increasing in i.

The elasticity of capital income can be written as

$$\begin{aligned}
\zeta_i^{a_R,(1-\tau_K)} &= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} + \int_{i'} \frac{dlog\,(a_i r_i)}{dlog\,(r_i)} \frac{dlog\,[r_i\,(\cdot)]}{dlog\,(a_{i'})} \frac{dlog\,(a_{i'})}{dlog\,(1-\tau_K)} di' \\
&= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} + \left(1 + \tilde{\zeta}_i^{a,r}\right) \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'. \qquad (18)
\end{aligned}$$

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

Comparative statics 1. This comparative statics compares the optimal general equilibrium capital tax to one in a self-confirming policy equilibrium. Firstly, express the capital gains elasticity in Equation (18) as

$$\zeta_i^{a_R,(1-\tau_K)} = \underbrace{\left(1+\varepsilon_i^{r,a}\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}}_{c_1} + \underbrace{\left(1+\varepsilon_i^{r,a} + \tilde{\zeta}_i^{a,r} \phi_i\right) \phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} \int_{i'} \delta_{i'}^{r,a} di'}_{c_2} \cdot \frac{1}{r_i}, \tag{19}$$

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

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

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

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

and

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

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

$$\frac{\tau_K^{GE}}{1-\tau_K^{GE}} = \frac{\tau_K^{PE}}{1-\tau_K^{PE}}\cdot\Delta + o\left(\tau_K^{GE}-\tau_K^{PE}\right), \tag{20}$$

where

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

Noting that $\Delta > 1$ since $\gamma_i^{r,(1-\tau_W)} < 0$ and $c_2 < 0$, as defined in Equation (18), concludes the proof.

Comparative statics 2. In this exercise, I compare the optimal capital gains tax in general equilibrium to the one in partial equilibrium holding all other primitives of the economy fixed. As in part (c), define $\bar{i}$ as the household who earns the average income $a_{R,\bar{i}} = \mathbb{E}(a_{R,i})$ and approximate each household's capital income around the income of $\bar{i}$ such that $\frac{a_{R,i}}{\mathbb{E}(a_{R,i})} = 1 + \zeta_i^{a_R,i} \frac{i-\bar{i}}{\bar{i}} + o(i-\bar{i})$. Then, notice that, in general equilibrium,

$$\begin{aligned}\zeta_i^{a,i} &= \tilde{\zeta}_i^{a,i} + \tilde{\zeta}_i^{a,r} \varepsilon_i^{r,a} \zeta_i^{a,i} + \tilde{\zeta}_i^{a,r} \int_{i'} \frac{dlog(r_i)}{dlog(a_{i'})} \frac{dlog(a_{i'})}{dlog(r_{i'})} \frac{dlog(r_{i'})}{dlog(a_i)} \zeta_i^{a,i} di \\ &= \tilde{\zeta}_i^{a,i} + \tilde{\zeta}_i^{a,r} \varepsilon_i^{r,a} \zeta_i^{a,i} + \tilde{\zeta}_i^{a,r} \int_{i'} \frac{1}{r_i} \delta_{i'}^{r,a} \tilde{\zeta}_{i'}^{a,r} \frac{1}{r_{i'}} \delta_i^{r,a} \zeta_i^{a,i} di\end{aligned}$$

Assuming that $\zeta_i^{a,i}$ is constant and using the fact that $\int_{i'} \delta_{i'}^{r,a} \frac{1}{r_{i'}} di = 0$, $\zeta_i^{a,i} = \tilde{\zeta}_i^{a,i} + \tilde{\zeta}_i^{a,r} \varepsilon_i^{r,a} \zeta_i^{a,i} = \phi_i \tilde{\zeta}_i^{a,i}$ which confirms the conjecture. Similarly,

$$\begin{aligned}\zeta_i^{a_R,i} &= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,i} + \tilde{\zeta}_i^{r,i} + \int_{i'} \frac{dlog(r_i)}{dlog(a_{i'})} \frac{dlog(a_{i'})}{dlog(r_{i'})} \frac{dlog(r_{i'})}{dlog(a_i)} \zeta_i^{a,i} di \\ &= \left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,i} + \tilde{\zeta}_i^{r,i} = \Phi_i \tilde{\zeta}_i^{a,i} + \tilde{\zeta}_i^{r,i}.\end{aligned}$$

Therefore, the relationship between household percentiles and their wealth and capital income remains unchanged because general equilibrium effects cancel out.

Then, one can approximate the capital income inequality in general equilibrium as

$$\begin{aligned}\mathbb{E}\left[\frac{\left(1 - \Gamma_i \left(1 + \gamma_i^{r,(1-\tau_K)}\right)\right) a_{R,i}}{\mathbb{E}(a_{R,i})}\right] &\approx \mathbb{E}\left[\left(1 - \Gamma_i \left(1 + \gamma_i^{r,(1-\tau_K)}\right)\right) \left(1 + \zeta_i^{a_R,i} \frac{i-\bar{i}}{\bar{i}}\right)\right] \\ &= \zeta_i^{a_R,i} \mathbb{E}\left[(1-\Gamma_i) \frac{i}{\bar{i}}\right] - \mathbb{E}\left[\Gamma_i \gamma_i^{r,(1-\tau_K)} \left(1 + \zeta_i^{a_R,i} \frac{i-\bar{i}}{\bar{i}}\right)\right],\end{aligned}$$

which is larger than $\zeta_i^{a_R,i} \mathbb{E}\left[(1-\Gamma_i) \frac{i}{\bar{i}}\right]$ since $\gamma_i^{r,(1-\tau_K)} < 0$. Therefore, capital income inequality is larger in general than in partial equilibrium. Using the above-derived relationships and observing that $\mathbb{E}\left(\delta_i^{r,a}\right) < 0$, the average capital income elasticity

$$\begin{aligned}\bar{\zeta}^{a_R,(1-\tau_K)} &= \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} \zeta_i^{a_R,(1-\tau_K)} di = \int_i \frac{a_{R,i}}{\mathbb{E}(a_{R,i})} \left[\left(1 + \varepsilon_i^{r,a}\right) \zeta_i^{a,(1-\tau_K)} + \left(1 + \tilde{\zeta}_i^{a,r}\right) \frac{1}{r_i} \int_{i'} \delta_{i'}^{r,a} \zeta_{i'}^{a,(1-\tau_K)} di'\right] di \\ &= \Phi_i \tilde{\zeta}_i^{a,(1-\tau_K)} \left(1 + \frac{1 + \tilde{\zeta}_i^{a,r}\left(1 + \Phi_i\right)}{1 + \varepsilon_i^{r,a}} \frac{\mathbb{E}(a_i)}{\mathbb{E}(a_{R,i})} \mathbb{E}\left(\delta_i^{r,a}\right)\right) < \Phi_i \tilde{\zeta}_i^{a,(1-\tau_K)}\end{aligned}$$

is revised downwards in general equilibrium. Altogether, $\tau_K^{GE} > \tau_K^{PE}$ holding all primitives other than the cross-return elasticities fixed.

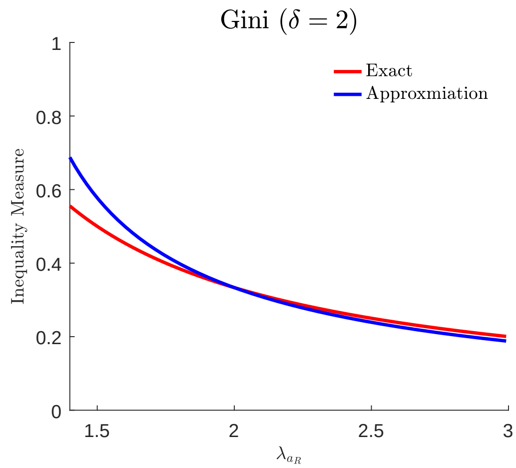

Figure 6: Approximation Error (Gini Social Welfare Function)

B Approximation Error

In Sections 2 and 4, I employ a first-order approximation of the measure of capital income inequality, allowing to remain agnostic about the distributions of wealth and return rates. The meaningfulness of this approach hinges on the performance of the approximation. Therefore, I now make specific distributional assumptions and compare the resulting exact measures of capital income inequality to their approximation. In the following approximations, I also account for the endogeneity of the household that earns the average capital income.

Following the literature on wealth and income inequality (see, for instance, Gabaix (2009)), suppose that capital income is Pareto distributed $\mathbb{P}(a_{R,i} \geq a_R) = a_{min}^{\lambda_{a_R}} a_R^{-\lambda_{a_R}}$. Observe that, under this assumption, $\zeta_i^{a_R,(1-i)} = -\frac{1}{\lambda_{a_R}}$. In the quantitative illustration of Section 4, I assume the Gini social welfare function, due to Sen (1974): $\mathit{\Gamma}_i = 2(1-i)$. This section extends the exposition to a more general class of rank-dependent social welfare functions $\mathit{\Gamma}_i = \frac{\delta}{\delta-1}\left(1 - i^{\delta-1}\right)$ where $\delta \geq 2$. This class is the so-called Lorenz or "A" family introduced by Aaberge (2000) that nests the Gini social welfare function ($\delta = 2$). Using these assumptions, one can derive the exact measure of capital income inequality. For $\delta = 2$, the inequality measure is exactly equal to the Gini coefficient $\mathbb{E}\left[(1-\mathit{\Gamma}_i)\frac{a_{R,i}}{\mathbb{E}(a_{R,i})}\right] = \frac{1}{2\lambda_{a_R}-1}$.

Figure 6 compares this exact expression to the proposed approximation for a range of values for λ_{a_R}. Overall the approximation performs well. Especially for a lower amount of inequality (larger λ_{a_R}), the approximation and the exact version of the Gini coefficient coincide. For a high level of inequality (small λ_{a_R}), the approximation

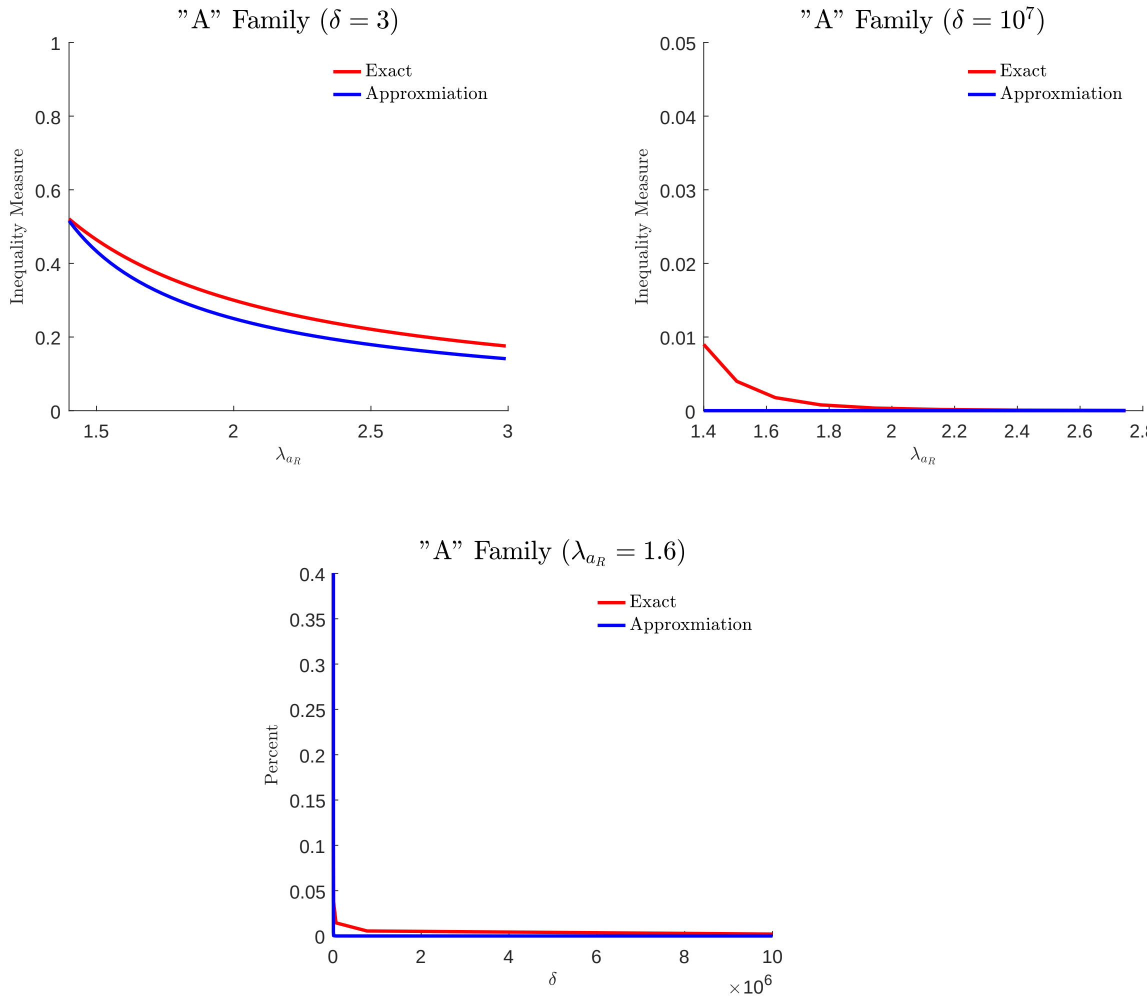

Figure 7: Approximation Error ("A" Family)

will overstate the Gini measure and, thus, the optimal tax rate. The approximation error is, however, limited (≈ 0.045) for the value used in Section 4 ($\lambda_{a_R} = 1.6$).

One can also study the approximation performance for other welfare functions. A higher value of δ means more weight on higher percentiles. In the limit ($\delta \to \infty$), the "A" family tends to the purely utilitarian case ($\Gamma_i = 1$). Figure 7 displays the approximation error for different values of δ. The upper two panels show the relationship between the inequality measures and the shape parameter for $\delta = 3$ and $\delta = 10^7$, respectively. Now, the approximation understates the measure of capital income inequality. However, the bias appears unsystematic over the range of shape parameters $\lambda_{a_R}$and is declining in δ. I further illustrate this observation in the lower panel by depicting the approximation error for different values of δ, holding the shape parameter fixed.

C A Dynamic Economy

In this section, I incorporate type and 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, $\tau_{W,t}$.[21] Suppose the initial distribution of $a_{i,0}$ is exogenously given.

Households can take effort $x_{i,t+1}$ at a cost $v(x_{i,t+1})$ to increase the rate of return $r'_{i,t}(x_{i,t}) > 0$ (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.[22] Intuitively, the higher an individual's hourly wage, the more she will work, and the more resources she can

[21] 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, $a_{R,i,t+1}(1 - \tau_{W,t+1}) = a_{i,t+1}[1 + (1 - \tau_{K,t+1})r_{t+1}]$ 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: $d\tau_{W,t+1} = d\tau_{K,i,t+1}\frac{r_{i,t+1}}{1+r_{i,t+1}} > 0$.

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

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 $r_{i,t} \equiv r_{i,t}(a_{i,t})$ where $r'_{i,t}(a_{i,t}) > 0$ and $r''_{i,t}(a_{i,t}) < 0$. When the costs are deductible from the tax base, define $r_{i,t} \equiv r_{i,t}(x_{i,t}) - v(x_{i,t})/a_{i,t}$. Return rates may also differ exogenously due to type dependence: $\frac{\partial r_{i,t}}{\partial i} \geq 0$. 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 $a_{R,i,t+1} \equiv a_{i,t+1}(1 + r_{i,t+1})$ and $\underline{a}_{i,t+1} \equiv a_{R,i,t+1}(1 - \tau_{W,t+1})$ are the pre- and after-tax final wealth. Altogether, households solve

$$\max_{c_{i,t}, l_{i,t}, a_{i,t+1}, x_{i,t+1}} U_{i,t}\left(c_{i,t}, a_{R,i,t+1}(1 - \tau_{W,t+1}), l_{i,t}, x_{i,t+1}\right) \tag{21}$$

subject to their budget constraint $c_{i,t} + a_{i,t+1} = a_{R,i,t}(1 - \tau_{W,t}) + w_{i,t}l_{i,t}(1 - \tau_{L,t}) + E_t$. 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 $\frac{\partial U_{i,t}(\cdot)}{\partial c_{i,t}} = \frac{\partial U_{i,t}(\cdot)}{\partial \underline{a}_{i,t+1}}(1 - \tau_{W,t+1})(1 + r_{i,t+1})$.

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 $\mathit{\Gamma}_{i,t} \geq 0$ as the Pareto weights. The government maximizes the sum of weighted utilities

$$\begin{aligned}\max_{\tau_W, \tau_L} \int_i & \mathit{\Gamma}_{i,t} U_{i,t}\left(a_{i,t}(1 + r_{i,t})(1 - \tau_W) + w_{i,t}l_{i,t}(1 - \tau_L) + E - a_{i,t+1},\right. \\ & \left. a_{i,t+1}(1 + r_{i,t+1})(1 - \tau_W), l_{i,t}\right) di\end{aligned} \tag{22}$$

subject to the balanced period budget $\tau_W a_{R,t} + \tau_L y_{L,t} = E$ and scale dependence $r_{i,t} \equiv r_{i,t}(a_{i,t})$. 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 $R \equiv \int_i r_{i,t}(a_{i,t})\, di$,

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

captures the reaction of savings with return heterogeneity, and

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

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 $r'_{i,t}(a_{i,t}) > 0$. 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, Corollary 1 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

$$a_{R,t} d\tau_W \left(1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}\right) = -y_{L,t} d\tau_L \left(1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}\right). \tag{23}$$

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

$$\int_i g_{i,t} \left[-\left(1 + \zeta_{i,t}^{a_R,(1-\tau_W)}\right) a_{R,i,t} d\tau_W + \frac{y_{L,i,t}}{y_{L,t}} \frac{1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}}{1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}} a_{R,t} d\tau_W - \frac{a_{i,t+1}}{1-\tau_W} d\tau_W \right] di = 0. \tag{24}$$

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 (24)

$$-\bar{a}^{initial} \left(1 + \hat{\zeta}^{a_R,(1-\tau_W)}\right) + \frac{1 - \zeta^{a_R,(1-\tau_W)} \frac{\tau_W}{1-\tau_W}}{1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}} \bar{y}_L - \frac{1}{1-\tau_W} \bar{a}^{final} = 0 \tag{25}$$

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 3 directly follows.

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

$$\tau_W = \frac{1 - \frac{\bar{a}^{initial}}{\bar{y}_L} \left(1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}\right) \left(1 + \hat{\zeta}^{a_R,(1-\tau_W)} + \frac{\bar{a}^{final}}{\bar{a}^{initial}}\right)}{1 + \zeta^{a_R,(1-\tau_W)} - \frac{\bar{a}^{initial}}{\bar{y}_L} \left(1 - \zeta^{y_L,(1-\tau_L)} \frac{\tau_L}{1-\tau_L}\right) \left(1 + \hat{\zeta}^{a_R,(1-\tau_W)}\right)} \tag{26}$$

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. Thus, I focus on the first comparative statics exercise in Proposition 1 (c), holding observables fixed. In this setting, this means fixing 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 ($g_{i,t}$). 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 ($\tilde{\zeta}_i^{a,r}$ 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, $\zeta^{a_R,(1-\tau_W)} > \zeta^{a_R,(1-\tau_W)}|_{\{r_i\}_{i\in[0,1]}}$ 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 ($\zeta^{a_R,(1-\tau_W)} > \zeta^{a_R,(1-\tau_W)}|_{\{r_i\}_{i\in[0,1]}}$). 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 3 is not in closed form. For small policy changes ($\tau_W \approx \tilde{\tau}_W$) from introducing a small amount of scale dependence ($\varepsilon_{i,t}^{r,a} \approx 0$),[23] one can use a first-order Taylor expansion to approximate aggregate wealth

$$a_{R,t}(\tau_W) = a_{R,t}(\tilde{\tau}_W)\left[1 + \frac{\tilde{\tau}_W - \tau_W}{1-\tilde{\tau}_W}\zeta^{a_R,(1-\tau_W)}\right] + o(\tau_W - \tilde{\tau}_W), \tag{27}$$

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.

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

Formally, $a_{R,t}(\tau_W) > a_{R,t}(\tilde{\tau}_W)$ for $\tilde{\tau}_W > \tau_W$.

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

$$\overline{a}^{final}(\tau_W) = \overline{a}^{final}(\tilde{\tau}_W)\frac{1+\frac{\tilde{\tau}_W-\tau_W}{1-\tilde{\tau}_W}\zeta^{a,(1-\tau_W)}}{1+\frac{\tilde{\tau}_W-\tau_W}{1-\tilde{\tau}_W}\zeta^{a_R,(1-\tau_W)}} + o(\tau_W-\tilde{\tau}_W). \tag{28}$$

If $\tilde{\tau}_W > \tau_W$, $\overline{a}^{final}(\tau_W) < \overline{a}^{final}(\tilde{\tau}_W)$ 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

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

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

$$\overline{a}^{initial}(\tau_W) = \overline{a}^{initial}(\tilde{\tau}_W) + o(\tau_W-\tilde{\tau}_W). \tag{29}$$

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 dependence as in the former the elasticity of capital is higher.[24]

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

$$\begin{aligned} SWF = \sum_{t=0}^{\infty}\beta^t \int_i \mathit{\Gamma}_{i,t}U_{i,t}\,(&a_{i,t}(1+r_{i,t})(1-\tau_{W,t}) + w_{i,t}l_{i,t}(1-\tau_{L,t}) + E_t - a_{i,t+1}, \\ &a_{i,t+1}(1+r_{i,t+1})(1-\tau_{W,t+1}), l_{i,t})\,di \end{aligned} \tag{30}$$

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 $r_{i,t} \equiv r_{i,t}(a_{i,t})$, 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

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

effect on social welfare, $dSWF = 0$, one obtains a dynamic version of the optimality condition from above

$$-\bar{a}^{initial}\left(1+(1-\beta)\sum_{t=T}^{\infty}\beta^{t-T}\zeta_t^{a_R,(1-\tau_W)}\right)+\bar{y}_L(1-\beta)\sum_{t=T}^{\infty}\beta^{t-T}\frac{1-\zeta_t^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta_t^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}}-\frac{1}{1-\tau_W}\frac{1}{\beta}\bar{a}^{final}=0. \tag{31}$$

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 $\gamma_{i,i',t}^{r,a}\equiv\frac{\partial log(r_{i,t})}{\partial log\left(a_{i',t}\right)}$.[25] 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 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 effects

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

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

[25] 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 $r_i(\cdot)$.

elasticities are adjusted. In the model of Section E with linear information costs, $\delta^{r,a}_{i',t}$ is positive for small values of $a_{i',t}$ and negative for large ones. This resembles a situation of trickle-up, in which cutting the top tax shifts economic rents from the bottom to the top.

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

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

and

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

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 effects, the optimal tax rate is defined by the optimality condition

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

which can be written as

$$-\,\overline{a}^{initial}\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}\right) + \frac{1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}}{1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}}\overline{y}_L - \frac{1}{1-\tau_W}\overline{a}^{final}\left(1+\hat{\gamma}^{1+r,(1-\tau_W)}\right) = 0 \tag{32}$$

using the notation from above and defining

$$\hat{\gamma}^{1+r,(1-\tau_W)} \equiv \int_i \left(1+\zeta^{a,1+r}_{i,t}\right)\left(\int_{i'} \gamma^{1+r,a}_{i,i',t+1}\zeta^{a,(1-\tau_W)}_{i',t+1}di'\right) g_{i,t}\frac{a_{i,t+1}}{a_{R,t}}di \Big/ \int_i g_{i,t}\frac{a_{i,t+1}}{a_{R,t}}di.$$

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 2), 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 4.

Proposition 4 (Optimal capital tax in general equilibrium)**.** *The optimal capital tax in the long-run steady-state general equilibrium is*

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

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 Proposition 2, 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 ($\tilde{\zeta}_i^{a,r}$ 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 (33). Again, a higher capital tax (e.g., $\tau_W^{GE} > \tau_W^{PE}$) reduces aggregate wealth (e.g., $a_{R,t}\left(\tau_W^{GE}\right) < a_{R,t}\left(\tau_W^{PE}\right)$)

$$a_{R,t}\left(\tau_W^{GE}\right) = a_{R,t}\left(\tau_W^{PE}\right)\left[1 + \frac{\tau_W^{PE} - \tau_W^{GE}}{1 - \tau_W^{PE}}\zeta^{a_R,(1-\tau_W)}\right] + o\left(\tau_W^{GE} - \tau_W^{PE}\right). \tag{34}$$

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 effect, $\hat{\gamma}^{1+r,(1-\tau_W)}$, 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.[26] 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 $\overline{a}^{final}_{t+1} = \beta (1-\tau_{W,t+1})\, \overline{a}^{initial}_{t+1}$ with the definitions from the deterministic version of the model $\overline{a}^{initial}_{t+1} \equiv \int_i g_{i,0} \frac{a_{R,i,t+1}}{a_{R,t+1}} di$ and $\overline{a}^{final}_{t+1} \equiv \int_i g_{i,0} \frac{a_{i,t+1}}{a_{R,t+1}} di$ and Pareto weights $\{\Gamma_{0,i}\}_{i \in [0,1]}$.

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 ($\Gamma_{0,i} = 1$) 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

$$SWF_\infty \equiv \sum_{t=0}^{\infty} \beta^t \mathbb{E} \left[u_{i,t}\left((1-\tau_W)\, a_{R,i,t} + (1-\tau_L)\, y_{L,i,t} + E - a_{i,t+1}, l_{i,t}\right)\right] \tag{35}$$

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

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

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

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

$$0 = -\sum_{t=0}^{\infty}\beta^t\mathbb{E}\left[\frac{\partial u_{i,0}\left(\cdot\right)}{\partial c_{i,0}}a_{R,i,0}\left(1+\zeta_i^{a_R,(1-\tau_W)}\right)\right]$$
$$-\sum_{t=0}^{\infty}\beta^t\mathbb{E}\left[\frac{\partial u_{i,t}\left(\cdot\right)}{\partial c_{i,t}}\frac{a_{i,t+1}}{1-\tau_W}+\frac{\partial u_{i,t}\left(\cdot\right)}{\partial c_{i,t}}a_{R,t}\frac{\left(1-\zeta^{a_R,(1-\tau_W)}\frac{\tau_W}{1-\tau_W}\right)}{\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)}\frac{y_{L,i,t}}{y_{L,t}}\right] \tag{36}$$

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

$$\tau_W = \frac{1-\frac{(1-\beta)\overline{a}^{initial}}{\overline{y}_L}\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}+\frac{\overline{a}^{final}}{(1-\beta)\overline{a}^{initial}}\right)}{1+\zeta^{a_R,(1-\tau_W)}-\frac{(1-\beta)\overline{a}^{initial}}{\overline{y}_L}\left(1-\zeta^{y_L,(1-\tau_W)}\frac{\tau_L}{1-\tau_L}\right)\left(1+\hat{\zeta}^{a_R,(1-\tau_W)}\right)} \tag{37}$$

with the only difference to Proposition 3 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

$$\zeta_{i,t}^{a,(1-\tau_W)} = \frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_{E,r_{i,t}} + \frac{dlog\left(a_{i,t}\right)}{dlog\left(r_{i,t}\right)}\frac{dlog\left[r_{i,t}\left(a_{i,t}\right)\right]}{dlog\left(a_{i,t}\right)}\frac{dlog\left(a_{i,t}\right)}{dlog\left(1-\tau_W\right)}|_E = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} \tag{38}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} = \frac{dlog\left[1+r_{i,t}\left(a_{i,t}\right)\right]}{dlog\left(1-\tau_W\right)} + \zeta_{i,t}^{a,(1-\tau_W)} = \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)}, \tag{39}$$

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 (23).

To obtain Equation (24), plug the households' first-order conditions $\frac{\partial U_{i,t}(\cdot)}{\partial c_{i,t}} = \frac{\partial U_{i,t}(\cdot)}{\partial \underline{a}_{i,t+1}} \frac{(1-\tau_{W,t+1})a_{R,i,t+1}}{a_{i,t+1}}$ and Equation (23) into

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

and set this expression equal to zero. Equation (25) follows from

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

Rearrange this equation to get the optimal wealth tax in Proposition 3.

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

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

Integrate out to get Equation (27)

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

Plug Equation (27) and

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

into

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

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 (28) follows.

Proceed along the same lines, to obtain Equation (29)

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

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 (23) 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 (31).

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 $\gamma_{i,i',t}^{r,a} = \frac{1}{r_{i,t}}\delta_{i'}^{r,a}$. One can simplify the second term on the right-hand side to

$$\begin{aligned}\int_{i'} \delta_{i'}^{r,a}\zeta_{i',t}^{a,(1-\tau_W)}di' &= \int_{i'} \delta_{i',t}^{r,a}\phi_{i',t}\tilde{\zeta}_{i',t}^{a,(1-\tau_W)}di' + \int_{i'} \delta_{i',t}^{r,a}\phi_{i',t}\zeta_{i',t}^{a,r}\frac{1}{r_{i',t}}di' \int_{i''} \delta_{i'',t}^{r,a}\tilde{\zeta}_{i'',t}^{a,(1-\tau_W)}di'' \\ &= \frac{1}{1-\int_{i'} \delta_{i',t}^{r,a}\phi_{i',t}\zeta_{i',t}^{a,r}\frac{1}{r_{i',t}}di' \int_{i''}} \int_{i'} \delta_{i',t}^{r,a}\phi_{i',t}\tilde{\zeta}_{i',t}^{a,(1-\tau_W)}di'.\end{aligned}$$

The wealth elasticity can be derived as

$$\begin{aligned}\zeta_{i,t}^{a_R,(1-\tau_W)} &= \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)} + \int_{i'} \frac{dlog\left(a_{i,t}\left(1+r_{i,t}\left(a_{i,t}\right)\right)\right)}{dlog\left(1+r_{i,t}\right)}\frac{dlog\left[1+r_{i,t}\left(\cdot\right)\right]}{dlog\left(a_{i',t}\right)}\frac{dlog\left(a_{i',t}\right)}{dlog\left(1-\tau_W\right)}|_E\, di' \\ &= \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)} + \left(1+\zeta_{i,t}^{a,1+r}\right)\int_{i'} \gamma_{i,i',t}^{1+r,a}\zeta_{i',t}^{a,(1-\tau_W)}di'.\end{aligned}$$

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

$$\zeta_{i,t}^{a,(1-\tau_W)} = \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}\left(1+\frac{1}{r_{i,t}}\phi_{i,t}\zeta_{i,t}^{a,r}\int_{i'} \delta_{i'}^{r,a}di'\right) < \phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} \tag{40}$$

and

$$\zeta_{i,t}^{a_R,(1-\tau_W)} = \left(1+\varepsilon_{i,t}^{1+r,a}\right)\zeta_{i,t}^{a,(1-\tau_W)} + \frac{1+\zeta_{i,t}^{a,1+r}}{1+r_{i,t}}\phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)}\int_{i'} \delta_{i',t}^{r,a}di' < \left(1+\varepsilon_{i,t}^{1+r,a}\right)\phi_{i,t}\tilde{\zeta}_{i,t}^{a,(1-\tau_W)} \tag{41}$$

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

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

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

Insert this equation and, as before, the households' first-order conditions and Equation (23) into

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

Set this expression equal to zero and use the definitions of the distributional parameters to get Equation (32). Equation (33) follows from rearranging Equation (32).

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 (34).

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 (scale dependence). I also show how to incorporate type dependence. 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.[27]

In each subperiod, there is a risk-free asset in infinitely elastic supply that delivers a return 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.[28]

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,

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

[28] 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.

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.[29]

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 $\mathbb{E}_{i,h}(\cdot|\mathcal{F}_{i,h})$ 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

$$r^p_{i,h} z = \varsigma_{i,h} \frac{\Pi_h - P_h}{P_h} + (1-\varsigma_{i,h})\, r^f_h z \tag{42}$$

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.

[29] Observe the implicit assumption that knowledge fully depreciates intertemporally. Any departure from this assumption would, just as non-convex cost functions, strengthens the main results.

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

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

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 $R_i \equiv \sum_{h=1}^{H} r_{i,h}^{p} z$ 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 can also be written as

$$H \cdot u\left[\left(1-\tau_W\right)\left(a_{i,1}\left(1 + r_{i,1}^{p} z\right) - v\left(x_{i,1}\right) z\right)\right] + o\left(1\right) \tag{43}$$

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*.[30] 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

$$\max_{x} \mathbb{E}_i\left(\max_{\varsigma} \mathbb{E}_i\left(u\left[\left(1-\tau_W\right)\left(a_i\left(1+r_i^{p} z\right) - v\left(x_i\right) z\right)\right] | \mathcal{F}_i\right)\right) \tag{44}$$

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

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

To get this expression, approximate the expected utility around the mean portfolio return. Thus, second-period utility features a deterministic mean-variance trade-off

[30] 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.

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 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

$$\varsigma_i = \varsigma\left(S_i, x_i, a_i; P, \mathcal{I}\right) \equiv arg\,\underset{\varsigma}{max}\ \mathbb{E}_i\left(u\left[\left(1-\tau_W\right)\left(a_i\left(1+r_i^p z\right) - v\left(x_i\right) z\right)\right] | \mathcal{F}_i\right) \tag{45}$$

and

$$x_i = x\left(a_i; \mathcal{I}\right) \equiv arg\,\underset{x}{max}\ \mathbb{E}_i\left[\underset{\varsigma}{max}\ \mathbb{E}_i\left(u\left[\left(1-\tau_W\right)\left(a_i\left(1+r_i^p z\right) - v\left(x_i\right) z\right)\right] | \mathcal{F}_i\right)\right], \tag{46}$$

(2) P clears the stock market

$$\int_i \frac{\varsigma_i a_i}{P} di = \theta, \tag{47}$$

and

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

individual information precision

$$\mathcal{I} = \frac{1}{\sigma^2} \int_i \frac{x\left(a_i; \mathcal{I}\right) \psi\left(a_i\right)}{h_0\left(\mathcal{I}\right) + x\left(a_i; \mathcal{I}\right)} di. \tag{48}$$

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

$$v'\left(x_i\right) = \frac{1}{2\rho} a_i \mathcal{S}'\left(x_i; \mathcal{I}\right), \tag{49}$$

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, $x\left(a_i; \mathcal{I}\right)$ 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 effect on an individual investor via the equilibrium price. Investors do not internalize this effect. 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:*

$$\mathbb{E}\left(r_i^{pe} z\right) = \mathbb{E}\left(r_i^{p} z\right) - r^f z = \frac{1}{\rho}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o\left(z\right), \tag{50}$$

$$\mathbb{V}\left(r_i^{pe} z\right) = \mathbb{V}\left(r_i^{p} z\right) = \frac{1}{\rho^2}\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z + o\left(z\right), \tag{51}$$

and

$$\frac{\mathbb{E}\left(r_i^{pe} z\right)}{\sqrt{\mathbb{V}\left(r_i^{pe} z\right)}} = \sqrt{\mathcal{S}\left(a_i, \{a_j\}_{j\in[0,1]}\right) z} + o\left(1\right). \tag{52}$$

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 $\mathbb{E}\left(\theta\right) = 0$ and $v\left(x_i\right) = \kappa x_i$. 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 $a_0 > a_i^*\left(\mathcal{I}\right)$. Then, the elasticity of the return (in a given subperiod) with respect to the amount of investment is positive

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

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 effects. 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 is analogous to 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., $a_0 > a_i^* (\mathcal{I})$). Suppose that $a_i = a_i^* (\mathcal{I})$ 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 $\mathbb{E}(\theta) = 0$, $\sigma^2 = \sigma_b^2 = \sigma_\theta^2 = 1$, and $v(x_i) = \kappa x_i$. Moreover, assume that $a_0 > a_i^* (\mathcal{I})$.

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 $v(x_i, y_i)$ 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.[31] As labor supply increases with i, this force amplifies the

[31] $\frac{\partial v(x_i,y_i)}{\partial x_i \partial y_i} < 0$ 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.

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 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

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

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

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

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

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

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

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

Therefore, one can rewrite the utility from final wealth as

$$\begin{aligned}u(a_{i,H+1}) &= u\left[(1 - \tau_W)\left(a_{i,1}\left(1 + \sum_{h=1}^{H} r_{i,h}^{p} z\right) - \sum_{h=1}^{H} v(x_{i,h}) z\right)\right] + o(z) \\ &= u\left[(1 - \tau_W)(a_{i,1}(1 + R_i) - v(X_i))\right] + o(z),\end{aligned}$$

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

$$
\begin{aligned}
u\left(a_{i,H+1}\right) &= u\left(a_{i,1}+\sum_{h=1}^{H}\Delta a_h\right)+o\left(z\right)=u\left(\Delta a_1,...,\Delta a_H\right)+o\left(z\right) \\
&= u\left(0,...,0\right)+\sum_{h=1}^{H}\frac{\partial u\left(0,...,0\right)}{\partial \Delta a_h}\Delta a_h+o\left(z\right)=u\left(a_{i,1}\right)+\sum_{h=1}^{H}\left(u\left(a_{i,1}+\Delta a_h\right)-u\left(a_{i,1}\right)\right)+o\left(z\right) \\
&= \sum_{h=1}^{H}u\left(a_{i,1}\left(1+r_{i,h}^{p}z\right)-v\left(x_{i,h}\right)z\right)+o\left(1\right),
\end{aligned}
$$

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 (43) 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

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

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 $\xi \equiv b-\frac{1}{\mathcal{I}}\theta$*

$$
log\left(P\right)=pz=\left(p_0+p_\xi\xi-r^f\right)z+o\left(z\right) \tag{53}
$$

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 $p_\xi \equiv 1-\frac{\mathcal{N}}{\mathcal{T}\sigma_b^2}$. The optimal investment in the risky *asset is given by*

$$\varsigma_i = \frac{1}{\rho\sigma\sqrt{z}}\lambda_i + o\left(1\right) \tag{54}$$

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 (53)), 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

$$\begin{aligned}0 &= \mathbb{E}_i\left[u'\left(\left(1-\tau_W\right)\left(a_i\left(1+\varsigma_i\frac{\Pi-P}{P}+\left(1-\varsigma_i\right)r^f z\right)-v\left(x_i\right)z\right)\right)\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right]\\ &= u'\left(\left(1-\tau_W\right)a_i\right)\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right]\\ &\quad + \left(1-\tau_W\right)a_i\varsigma_i u''\left(\left(1-\tau_W\right)a_i\right)\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)^2|\mathcal{F}_i\right] + o\left(z\right)\end{aligned} \tag{55}$$

that determines the optimal portfolio choice. Note that

$$\begin{aligned}\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)|\mathcal{F}_i\right] &= \mathbb{E}_i\left[\left(\frac{exp\left(\pi z\right)-exp\left(pz\right)}{exp\left(pz\right)}-r^f z\right)|\mathcal{F}_i\right]\\ &= \mathbb{E}_i\left[\left(\frac{1+\pi z+\frac{1}{2}\left(\pi z\right)^2+o\left(z^2\right)-1-pz-\frac{1}{2}\left(pz\right)^2-o\left(z^2\right)}{1+pz+o\left(z\right)}\right)|\mathcal{F}_i\right]-r^f z\\ &= \mathbb{E}_i\left(\pi z|\mathcal{F}_i\right)+\frac{1}{2}\mathbb{E}_i\left[\left(\left(\pi z\right)^2-\left(pz\right)^2\right)|\mathcal{F}_i\right]-pz-r^f z+o\left(z\right)\\ &= \mathbb{E}_i\left(bz|\mathcal{F}_i\right)+\frac{1}{2}\sigma^2 z-pz-r^f z+o\left(z\right)\end{aligned} \tag{56}$$

and

$$\mathbb{E}_i\left[\left(\frac{\Pi-P}{P}-r^f z\right)^2|\mathcal{F}_i\right]=\mathbb{E}_i\left[\left(\pi z+\frac{1}{2}(\pi z)^2-pz-\frac{1}{2}(pz)^2-r^f z\right)^2|\mathcal{F}_i\right]+o(z)$$
$$=\mathbb{E}_i\left[(\pi z)^2|\mathcal{F}_i\right]+o(z)=\sigma^2 z+o(z). \tag{57}$$

Plug these expressions and the conjectured equilibrium price into Equation (55). To get Equation (54), 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 (54) and the definitions of the aggregate variables into the stock market clearing condition (Equation (45)) 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 (53) 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*

$$v'(x_i)=\frac{a_i}{2\rho}\mathcal{S}'(x_i;\mathcal{I}), \tag{58}$$

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

$$\mathcal{I}=\frac{1}{\rho\sigma^2}\int_{a_i^*(\mathcal{I})}^{a_1}\frac{a_i x_i(a_i;\mathcal{I})}{h_0(\mathcal{I})+x_i(a_i;\mathcal{I})}dG(a_i). \tag{59}$$

There exists a unique log-linear equilibrium.

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

$$\mathcal{S}(x_{i,1};\mathcal{I})z\equiv\mathbb{E}_i\left(\lambda_i^2\right)=\mathbb{V}_i(\lambda_i)+\mathbb{E}_i(\lambda_i)^2$$
$$=-\frac{z}{\sigma^2}\frac{1}{h_0(\mathcal{I})+x_i}+\frac{z}{\sigma^2}\left[\frac{\sigma_\theta^2}{\mathcal{I}^2}p_\xi^2+\sigma_b^2(1-p_\xi)^2+\frac{E(\theta)^2}{\mathcal{I}^2}\left(1-h_0(\mathcal{I})\frac{\mathcal{N}}{\mathcal{T}}\right)^2\right]. \tag{60}$$

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 (56) and (57) and the definition in Equation (60), by the law of total expectation, expected returns read as

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

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 (50) and (51). Equation (52) follows from dividing (50) by the square root of (51). 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 (49) 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 (50) and (60)

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

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

$$\varepsilon_i^{\mathbb{E}(r^pz),a}\equiv\frac{a_i}{\mathbb{E}\left(r_i^pz\right)}\frac{d\mathbb{E}\left(r_i^pz\right)}{da_i}=\frac{1}{\mathbb{E}\left(r_i^pz\right)}\frac{z}{\sigma}\sqrt{\frac{\kappa}{2\rho a_i}}=\frac{\sqrt{\kappa/\left(2\rho\sigma^2a_i\right)}}{\mathcal{S}\left(a_i,\left\{a_j\right\}_{j\in[0,1]}\right)/\rho+r^f}. \tag{61}$$

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

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

where $\delta_{i'}^{\mathbb{E}(r^pz),a}$ is decreasing in $a_{i'}$ and $\delta_{i'}^{\mathbb{E}(r^pz),a}\geq0$ for $a_{i'}\leq\tilde{a}$. Then, $\delta_{i'}^{\mathbb{E}(r^pz),a}<0$ 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

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

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

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

$$\begin{aligned}
\delta_{i'}^{\mathbb{E}(r^p z),a} &= a_{i'} \frac{2z}{\rho \mathcal{I}} \left(\frac{1}{\mathcal{T}^2} \frac{\mathcal{N}}{\mathcal{T}} \cdot \frac{\partial \mathcal{T}_{i'}}{\partial a_{i'}} - \frac{1}{\mathcal{T}^2} \cdot \frac{\partial \mathcal{N}_{i'}}{\partial a_{i'}} - \frac{1}{\mathcal{I}^2} \frac{\partial \mathcal{I}_{i'}}{\partial a_{i'}} \right) \\
&= a_{i'} \frac{2z}{\rho \mathcal{I}} \left[\left(\frac{1}{\mathcal{T}^2} \frac{\mathcal{N}}{\mathcal{T}} - \frac{\left(1 - \mathcal{N}/\mathcal{T}\right)^2}{\mathcal{I}^2 \left(1 + 2\mathcal{I}\mathcal{N}\right)} \right) \cdot \frac{\partial \mathcal{T}_{i'}}{\partial a_{i'}} + \left(\frac{\left(1 - \mathcal{N}/\mathcal{T}\right)^2}{\mathcal{I}^2 \left(1 + 2\mathcal{I}\mathcal{N}\right)} \left(1 + \mathcal{I}^2\right) - \frac{1}{\mathcal{T}^2} \right) \cdot \frac{\partial \mathcal{N}_{i'}}{\partial a_{i'}} \right] \\
&= \frac{2z a_{i'}}{\rho \mathcal{T}^2 \mathcal{I}^3 \left(1 + 2\mathcal{I}\mathcal{N}\right)} \left[\Omega_{\mathcal{T}} \cdot \frac{1}{\rho} + \Omega_{\mathcal{N}} \cdot \sqrt{\kappa / \left(2 a_{i'} \rho\right)} \right], \qquad (62)
\end{aligned}$$

where

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

respectively. For $T_k'\left(0\right)=0$, 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 with scale and type 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 first 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

$$u\left(l_i, a_i, e_i; w_i\right) \equiv u_0\left(c_{i,0} - v_0\left(l_i\right)\right) + \beta u_1\left(c_{i,1} - v_1\left(x_i\right)\right), \tag{63}$$

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 ($\frac{\partial r_i}{\partial a_i} > 0$) holds in this economy (for $T_k(0) = 0$, $T_k'(0) = 0$, and $T_k''(0) = 0$) irrespective of this deductibility. Therefore, return rates exhibit scale dependence. Moreover, let $\frac{\partial r_i}{\partial i} \geq 0$ so that there may also be type dependence. 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

$$c_{i,0} + a_i \leq w_i l_i - T_l \left(w_i l_i \right). \tag{64}$$

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

$$c_{i,1} \leq a_i \left(1 + r_i \right) - T_k \left(r_i a_i \right). \tag{65}$$

Let $\mathcal{U}_i \left(T_l, T_k \right)$ denote household i's indirect utility from optimally choosing savings, a_i, and effort levels, $\{ l_i, x_i \}$, to maximize Equation (63) subject to Equations (64) and (65). 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

$$\mathcal{R} \left(T_l, T_k \right) \equiv \int_0^1 T_l \left(w_i l_i \right) di + \beta \int_0^1 T_k \left(r_i a_i \right) di \geq \bar{E}. \tag{66}$$

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

$$\mathcal{G} \left(T_l, T_k \right) \equiv \int_0^1 \mathit{\Gamma}_i \mathcal{U}_i \left(T_l, T_k \right) di \tag{67}$$

subject to Equation (66), where $\mathit{\Gamma}_i$ denotes household i's Pareto weight with $\int_0^1 \mathit{\Gamma}_i di = 1$. Denote λ as the marginal value of public funds and $g_{i,t} \equiv \left(1/\lambda \right) \mathit{\Gamma}_i u_t' \left(c_{i,t} - v_t \left(\cdot \right) \right)$ 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 U.S. 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 (66) and (67) 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 $I_{0,i}\equiv y_iT_l'\left(y_i\right)-T_l\left(y_i\right)$ and $I_{1,i}\equiv a_{R,i}T_k'\left(a_{R,i}\right)-T_k\left(a_{R,i}\right)$ as the virtual income of individual i in period 0 and period 1, respectively. Define $\zeta_i^{a,\left(1-T_t'\right)}$ (η_i^{a,I_t}) 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 ($\zeta_i^{l,\left(1-T_k'\right)}=\eta_i^{l,I_1}=0$). Moreover, let $\tilde{\zeta}_i^{a,r}$ be the elasticity of savings with respect to the rate of return.[32]

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

[32]Similar to $\tilde{\zeta}_i^{a,\left(1-T_k'\right)}$ and $\tilde{\eta}_i^{a,I_k}$, the definition of $\tilde{\zeta}_i^{a,r}$ 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.

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

$$\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)}{a_i} = -\underbrace{\tilde{\zeta}_i^{a,(1-T_k')}}_{>0}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \underbrace{\tilde{\eta}_i^{a,I_2}}_{\leq 0}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)} + \underbrace{\tilde{\zeta}_i^{a,r}}_{>0}\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)}{r_i}. \tag{68}$$

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

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{PE}}{r_i} = \varepsilon_i^{r,a}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{PE}}{a_i}. \tag{69}$$

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

$$\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{GE}}{r_i} = \varepsilon_i^{r,a}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{GE}}{a_i} + \int_{i'}\gamma_{i,i'}^{r,a}\frac{\hat{a}_{i'}\left(T_k,\hat{T}_k\right)^{GE}}{a_{i'}}di'. \tag{70}$$

Thus, in both cases, one needs to upward adjust income and substitution effects of savings by an inequality multiplier effect $\phi_i \equiv \frac{1}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}} > 1$. 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 (68) and (70), 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 $\phi_i \equiv \frac{1}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}}$. In partial equilibrium, the* *first-order change in the optimal savings is given by*

$$\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{PE}}{a_i} = -\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \phi_i\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)}. \tag{71}$$

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

$$\begin{aligned}\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{GE}}{a_i} = &-\phi_i\tilde{\zeta}_i^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \phi_i\tilde{\eta}_i^{a,I_2}\frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)}\\ &-\phi_i\tilde{\zeta}_i^{a,r}\int_{i'}\phi_{i'}R_{i,i'}\left[\tilde{\zeta}_{i'}^{a,(1-T_k')}\frac{\hat{T}_k'\left(a_{R,i'}\right)}{1-T_k'\left(a_{R,i'}\right)} + \tilde{\eta}_{i'}^{a,I_2}\frac{\hat{T}_k\left(a_{R,i'}\right)}{a_{R,i'}\left(1-T_k'\left(a_{R,i'}\right)\right)}\right]di',\end{aligned} \tag{72}$$

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 $\phi_i = \frac{1}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}} = \sum_{n=0}^{\infty}\left(\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}\right)^n$. 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, $\mathcal{K}_{i,i'}^{(1)}$, describes the impact of savings by household i' on the returns of i. The second kernel $\mathcal{K}_{i,i'}^{(2)} = \int_{i''}\gamma_{i,i''}^{r,a}\phi_{i''}\zeta_{i''}^{a,r}\gamma_{i'',i'}^{r,a}di'$ 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 $\gamma_{i,i'}^{r,a}$ along the wealth distribution. Suppose, for instance, that $\gamma_{i,i'}^{r,a} > 0$ 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 $\gamma_{i,i'}^{r,a} \to 0$.

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

$$\frac{\hat{r}_i\left(T_k, \hat{T}_k\right)^{GE}}{r_i} = -\phi_i \varepsilon_i^{r,a} \tilde{\zeta}_i^{a,\left(1-T_k'\right)} \frac{\hat{T}_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)} - \phi_i \varepsilon_i^{r,a} \tilde{\eta}_i^{a,I_2} \frac{\hat{T}_k\left(a_{R,i}\right)}{a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)} + \phi_i CE_i, \tag{73}$$

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

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

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

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

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, Corollary 1.

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

$$\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right)^{PE} = -\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right) \tag{74}$$

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,

$$\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right)^{GE} = -\lambda g_{i,1}\left(\beta/\Gamma_i\right)\hat{T}_k\left(a_{R,i}\right) + \lambda g_{i,1}\left(\beta/\Gamma_i\right)a_{R,i}\left(1-T_k'\left(a_{R,i}\right)\right)\left(1+\tilde{\zeta}_i^{a,r}\right)CE_i. \tag{75}$$

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* $\int_{i'}\int_i\left|\phi_i\tilde{\zeta}_i^{a,r}\gamma_{i,i'}^{r,a}\right|^2 didi' < 1$*. 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*

$$\hat{\mathcal{G}}\left(T_k,\hat{T}_k\right)^{EQ} = \int_i \Gamma_i\hat{\mathcal{U}}_i\left(T_k,\hat{T}_k\right)^{EQ} di. \tag{76}$$

The first-order change in government revenues is given by

$$\hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{EQ}=\beta\int_i \hat{T}_k\left(a_{R,i}\right)di+\beta\int_i T_k'\left(a_{R,i}\right)a_{R,i}\left[\frac{\hat{r}_i\left(T_k,\hat{T}_k\right)^{EQ}}{r_i}+\frac{\hat{a}_i\left(T_k,\hat{T}_k\right)^{EQ}}{a_i}\right]di. \tag{77}$$

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 general equilibrium effects 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 $\hat{T}_k$:

$$\frac{1}{\lambda}\hat{\mathcal{G}}\left(T_k,\hat{T}_k\right)^{EQ}+\hat{\mathcal{R}}\left(T_k,\hat{T}_k\right)^{EQ}=0. \tag{78}$$

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 5 characterizes the optimal nonlinear capital gains tax.

Proposition 5 (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*

$$\begin{aligned}\frac{T_k'\left(a_{R,i}\right)^{PE}}{1-T_k'\left(a_{R,i}\right)^{PE}}=&\frac{1}{\zeta_i^{a_R,\left(1-T_k'\right)}}\frac{1-H\left(a_{R,i}\right)}{a_{R,i}h\left(a_{R,i}\right)}\\&\times\int_{a_{R,i}}^{\infty}\left(1-g_{i'',1}\right)exp\left[-\int_{a_{R,i}}^{a_{R,i'''}}\frac{\eta_{i'}^{a_R,T_k}}{\zeta_{i'}^{a_R,\left(1-T_k'\right)}}\frac{da_{R,i'}}{a_{R,i'}}\right]\frac{dH\left(a_{R,i''}\right)}{1-H\left(a_{R,i}\right)},\end{aligned}\tag{79}$$

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

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 $\mathit{\Phi}_i$ 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 $\mathit{\Phi}_i>1$), 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 (79) in terms of primitives:

$$\frac{T_k'\left(a_{R,i}\right)^{PE}}{1-T_k'\left(a_{R,i}\right)^{PE}} = \frac{\tilde{\zeta}_i^{a,I_0}\tilde{\zeta}_i^{I_0,i}+\tilde{\zeta}_i^{a,r}\tilde{\zeta}_i^{r,i}+\tilde{\zeta}_i^{r,i}/\Phi_i}{\tilde{\zeta}_i^{a,\left(1-T_k'\right)}}\frac{1-i}{i}$$
$$\times\int_i^1\left(1-g_{i'',1}\right)exp\left[-\int_i^{i''}\frac{\tilde{\eta}_{i'}^{a,T_k}}{\tilde{\zeta}_{i'}^{a,\left(1-T_k'\right)}}\frac{di'}{i'}\right]\frac{di''}{1-i}. \tag{80}$$

Therefore, when investment rates are endogenously determined (scale dependence), the capital gains tax is the same as the one without scale dependence holding all the other primitives of the economy fixed and in the absence of type dependence ($\tilde{\zeta}_i^{r,i}=0$). The upward adjustment in the savings elasticity just offsets the rise in observed inequality. When there is type dependence ($\tilde{\zeta}_i^{r,i}>0$), conditional on all other primitives, the capital gains tax is lower with than without scale dependence. The resulting adjustment depends on the relative strength of type and scale dependence $\tilde{\zeta}_i^{r,i}/\Phi_i$.

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 $\zeta^{a_R,\left(1-T_k'\right)}=\Phi\tilde{\zeta}^{a,\left(1-T_k'\right)}$, there is no type dependence $\tilde{\zeta}_i^{r,i}=0$ at the top and that there are no income effects $\tilde{\eta}=0$. Suppose that, without scale 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 $a_k=\tilde{a}_k/\Phi$. 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 U.S.) have not increased despite the drastic rise in top 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 6 describes the optimal capital tax.

Proposition 6 (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*

$$\frac{T_k'(a_{R,i})^{GE}}{1-T_k'(a_{R,i})^{GE}} = \frac{1}{\zeta_i^{a_R,(1-T_k')}}\frac{1-H(a_{R,i})}{a_{R,i}h(a_{R,i})} - \frac{\delta_i^{r,a}}{r_i\left(1+\varepsilon_i^{r,a}\right)(1+\Psi)}$$
$$\times \int_{\mathbb{R}_+} \frac{1+\left(1+\zeta_{i'}^{r,a}\right)\zeta_{i'}^{a,r}}{\zeta_{i'}^{a_R,(1-T_k')}}\frac{1-H(a_{R,i'})}{a_{R,i'}h(a_{R,i'})}\frac{a_{i'}\left[1-T_k'(a_{R,i'})^{GE}\right]}{a_i\left[1-T_k'(a_{R,i})^{GE}\right]}dH(a_{R,i'}), \tag{81}$$

where $\zeta_i^{a_R,(1-T_k')} \equiv \Phi_i\tilde{\zeta}_i^{a,(1-T_k')}$, $\Phi_i \equiv \left(1+\varepsilon_i^{r,a}\right)\phi_i$, $\phi_i \equiv \frac{1}{1-\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}}$, *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 (79)). 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.[33]

[33] 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, $\overline{T_k}'(a_{R,i})^{SCPE}$.

Then, one can write the general equilibrium tax rate as

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

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 2).

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

$$\frac{d}{d\mathcal{P}}\left[\frac{1}{\lambda}\mathcal{G}(T_l, T_k) + \mathcal{R}(T_l, T_k)\right] = \beta C'(\mathcal{P}). \tag{82}$$

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

$$\underbrace{\frac{d}{d\mathcal{P}}\int_i (\Gamma_i/\lambda)\left[u_0(\cdot) + \beta H u_1[\mathbb{E}(\cdot)] + \frac{1}{2}\beta H u_1''(a_i)\mathbb{V}(a_i r_i^p z)\right] di + o(z)}_{\equiv \mathcal{WE}_{\mathcal{P}}}$$
$$+ \underbrace{\frac{d}{d\mathcal{P}}\int_i \beta H T_k[a_i \mathbb{E}(r_i^p z)]\, di + o(z)}_{\equiv \mathcal{RE}_{\mathcal{P}}} = \beta C'(\mathcal{P}). \tag{83}$$

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

$$\mathcal{WE}_{\kappa} = \frac{1}{\kappa}\beta H \int_i g_{i,1}\left(\mathbb{E}\left(\cdot\right)\right) x_i z di, \tag{84}$$

whereas the revenue differential includes behavioral effects

$$\begin{aligned}\mathcal{RE}_{\kappa} = & \frac{1}{\kappa}\beta H \int_i \frac{T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]}{1-T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right]}\left(1+\varepsilon_i^{r,a}\right)\eta_i^{a,I_2} v\left(x_i\right) z di \\ & - \frac{1}{\kappa}\beta H \int_i T_k'\left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right)\left(1+\tilde{\zeta}_i^{a,r}\right)\zeta_i^{\mathbb{E}(r^p z),\kappa} di\end{aligned} \tag{85}$$

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 $v\left(x_i\right) = \kappa z \cdot max\left\{0, x_i - \underline{x}\right\}$. 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

$$\mathcal{WE}_{\underline{x}} = \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \int_0^{\underline{i}} \mathbb{E}\left[g_{i,1}\left(\cdot\right)\right] \frac{dlog\left[\mathbb{E}\left(u_1'\left(\cdot\right)\right)\right]}{dlog\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]} di + \frac{1}{\underline{x}}\beta H v\left(\underline{x}\right) z \int_{\underline{i}}^1 g_{i,1}\left[\mathbb{E}\left(\cdot\right)\right] di + o\left(z\right) \tag{86}$$

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

$$\begin{aligned}\mathcal{RE}_{\underline{x}} =& \frac{1}{\underline{x}}\beta H \int_0^{\underline{i}} T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right] a_i \mathbb{E}\left(r_i^p z\right) \left(1+\tilde{\zeta}_i^{a,r}\right) \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} di \\ &+ \frac{1}{\underline{x}}\beta H \int_i \frac{T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right]}{1-T_k' \left[\mathbb{E}\left(a_i r_i^p z\right)\right]} a_i \mathbb{E}\left(r_i^p z\right) \left(1+\varepsilon_i^{r,a}\right) \eta_i^{a,I_2} v\left(\underline{x}\right) z di + o\left(z\right) \end{aligned} \tag{87}$$

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, $C'\left(\mathcal{P}\right)$, 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

$$\begin{aligned} \left[l_i\right]:\ & 0 = w_i\left(1-T_l'\left(w_i l_i\right)\right) - v_0'\left(l_i\right) \\ \left[a_i\right]:\ & 0 = \left[1+r_i\left(1-T_k'\left(r_i a_i\right)\right)\right]\beta u_1'\left(\cdot\right) - u_0'\left(\cdot\right) \\ \left[e_i\right]:\ & 0 = a_i\left(1-T_k'\left(r_i a_i\right)\right) r_i'\left(e_i\right) - v_1'\left(e_i\right), \end{aligned} \tag{88}$$

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

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

where I use the definition of the local rate of tax progressivity $p_t\left(y\right) \equiv \frac{y T_t''(y)}{1 - T_t'(y)}$ for $t \in \{l, k\}$ and the assumption that $p_l\left(y_i\right) < 1$. 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

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

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

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

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

H.2 Incidence of Nonlinear Tax Reforms

Incidence on savings in partial equilibrium. To derive the incidence on savings in partial equilibrium, plug Equation (69) into (68)

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

Rearrange this expression to obtain Equation (71) in Lemma 4.

Incidence on savings in general equilibrium. To derive Equation (72) in Lemma 4, insert Equation (70) into (68) and rearrange

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

This expression is a Fredholm integral equation of the second kind. Suppose that $\int_{i'}\int_i\left|\phi_i\tilde{\zeta}_i^{a,r}\gamma_{i,i'}^{r,a}\right|^2 di di' < 1$. Then, by Theorem 2.3.1 in Zemyan (2012), the unique solution to this expression is given by Equation (72).

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 (69) and (71). 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

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

Incidence on revenues and welfare. Equation (76) is standard. Perturb Equation (66)

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

and rearrange to get (77).

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

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

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

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

This expression is a first-order linear differential equation. Use standard techniques (see Saez (2001)) to obtain Equation (79).

To express (79) in terms of the pre-tax wage distribution, change the variables in the integration

$$\frac{T_k'\left(a_{R,i}\right)}{1-T_k'\left(a_{R,i}\right)}=\frac{1}{\zeta_i^{a_R,\left(1-T_k'\right)}}\frac{1-i}{a_{R,i}h\left(a_{R,i}\right)}\int_i^1\left(1-g_{i'',1}\right)exp\left[-\int_i^{i''}\frac{\tilde{\eta}_{i'}^{a,T_k}}{\tilde{\zeta}_{i'}^{a,\left(1-T_k'\right)}}\frac{di'}{i'}\right]\frac{di''}{1-i}. \quad (89)$$

Since $i=H\left(a_{R,i}\right)$,

$$i=a_{R,i}h\left(a_{R,i}\right)\frac{da_{R,i}}{di}\frac{i}{a_{R,i}}. \quad (90)$$

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

$$\frac{da_{R,i}}{di}\frac{i}{a_{R,i}}=\left(1+\varepsilon_i^{r,a}\right)\frac{da_i}{di}\frac{i}{a_i}+\frac{\partial r_i\left(a_{R,i}\right)}{\partial i}\frac{i}{r_i\left(a_{R,i}\right)}\equiv\left(1+\varepsilon_i^{r,a}\right)\phi_i\tilde{\zeta}_i^{a,i}+\tilde{\zeta}_i^{r,i} \quad (91)$$

where the second equality follows from the fact that

$$\zeta_i^{a,i}=\tilde{\zeta}_i^{a,i}+\tilde{\zeta}_i^{a,r}\varepsilon_i^{r,a}\zeta_i^{a,i}=\phi_i\tilde{\zeta}_i^{a,i}=\phi_i\left(\tilde{\zeta}_i^{a,y-T_l(y)}\tilde{\zeta}_i^{y-T_l(y),i}+\tilde{\zeta}_i^{a,r}\tilde{\zeta}_i^{r,i}\right).$$

Plug Equations (90) and (91) into (89), to get Equation (80).

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

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

noting that

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

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

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

and the incidence on utility is given by

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

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

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

if and only if

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

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

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

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

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

$$= \frac{\int_{a_{R,i}} \frac{1+\left(1+\varepsilon_i^{r,a}\right)\tilde{\zeta}_i^{a,r}}{1+\varepsilon_i^{r,a}} \frac{1}{r_i}\delta_i^{r,a} dH\left(a_{R,i}\right) + 1 - \int_{a_{R,i}} \tilde{\zeta}_i^{a,r} \frac{1}{r_i}\delta_i^{r,a} dH\left(a_{R,i}\right)}{1-\int_{a_{R,i}} \tilde{\zeta}_i^{a,r}\frac{1}{r_i}\delta_i^{r,a} dH\left(a_{R,i}\right)}$$

$$\times \int_{a_{R,i}} a_i T_k'\left(a_{R,i}\right)\left[1+\left(1+\zeta_i^{r,a}\right)\tilde{\zeta}_i^{a,r}\right] dH\left(a_{R,i}\right)$$

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

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

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

which concludes the proof of Equation (81).

To compare this capital gains tax to the one by the exogenous technology planner, note that in the latter case

$$\frac{\overline{T_k}'\left(a_{R,i}\right)}{1-\overline{T_k}'\left(a_{R,i}\right)} = \frac{1}{\zeta_i^{a_{R},\left(1-T_k'\right)}} \frac{1-H\left(a_{R,i}\right)}{a_{R,i}h\left(a_{R,i}\right)}$$

and insert this expression into (81).

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

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

A household's tax liability can be approximated by

$$T_k\left(R_i a_i\right)=T_k\left(a_i r_{i,1} z+\ldots+a_i r_{i,H} z\right)+o\left(z\right)\equiv T_k\left(a_i r_{i,1} z,\ldots,a_i r_{i,H} z\right)+o\left(z\right)=\sum_{h=1}^{H} T_k'\left(0\right) a_i r_{i,h} z+o\left(z\right)$$
$$=T_k\left(a_i r_{i,1} z\right)-T_k\left(0\right)+\ldots+T_k\left(a_i r_{i,H} z\right)-T_k\left(0\right)+o\left(z\right)=\sum_{h=1}^{H} T_k\left(a_i r_{i,h} z\right)+o\left(z\right)$$

Using this expression, the first-order effect on revenues reads as

$$\begin{aligned}\mathcal{RE}_{\mathcal{P}}&\equiv\frac{d}{d\mathcal{P}}\mathcal{R}\left(T_l,T_k\right)=\frac{d}{d\mathcal{P}}\int_i \beta\mathbb{E}\left(T_k\left[R_i a_i\right]\right)di=\frac{d}{d\mathcal{P}}\int_i \beta\sum_{h=1}^{H} T_k\left[a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right]di+o\left(z\right)\\&+\frac{d}{d\mathcal{P}}\int_i \beta\sum_{h=1}^{H} T_k'\left[a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right]\mathbb{E}\left[a_i r_{i,h}^{p} z-a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right]di\\&+\frac{1}{2}\frac{d}{d\mathcal{P}}\int_i \beta\sum_{h=1}^{H} T_k''\left[a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right]\mathbb{E}\left[\left(a_i r_{i,h}^{p} z-a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right)^2\right]di+o\left(z\right)\\&=\frac{d}{d\mathcal{P}}\int_i \beta\sum_{h=1}^{H} T_k\left[a_i\mathbb{E}\left(r_{i,h}^{p} z\right)\right]di+o\left(z\right)=\frac{d}{d\mathcal{P}}\int_i \beta H T_k\left[a_i\mathbb{E}\left(r_i^{p} z\right)\right]di+o\left(z\right)\end{aligned}$$

since, in partial equilibrium,

$$\begin{aligned}\mathbb{E}\left(r_{i,h}^{p} z\right)&=\frac{1}{\rho}\mathcal{S}\left(a_{i,h}\right)z+r^f z+o\left(z\right)=\frac{1}{\rho}\mathcal{S}\left(a_i\Pi_{s=1}^{h}\left(1+r_{i,h}^{p} z\right)\right)z+r^f z+o\left(z\right)\\&=\frac{1}{\rho}\mathcal{S}\left(a_i+a_i\sum_{s=1}^{h} r_{i,h}^{p} z\right)z+r^f z+o\left(z\right)=\frac{1}{\rho}\mathcal{S}\left(a_i\right)z+r^f z+o\left(z\right)=\mathbb{E}\left(r_i^{p} z\right).\end{aligned}$$

Cost subsidy. For a cost subsidy, $\mathcal{P}=\Delta\kappa<0$, the first-order welfare effect can be written as in Equation (84)

$$\mathcal{WE}_{\kappa}=\int_i \left(\varGamma_i/\lambda\right)\beta H u_1'\left[\mathbb{E}\left(\cdot\right)\right]x_i z di\equiv\frac{1}{\kappa}\beta H\int_i g_{i,1}\left(\mathbb{E}\left(\cdot\right)\right)v\left(x_i\right)z di+o\left(z\right)$$

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 (85)

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

Financial education. When the government provides a minimal level of financial knowledge, $\underline{x}$, for free, such that the information cost reads as $v(x_i) = \kappa z \cdot max\{0, x_i - \underline{x}\}$, 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 $\underline{i}$

$$\begin{aligned}
\mathcal{WE}_{\underline{x}} &= \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \int_0^{\underline{i}} \left(\mathit{\Gamma}_i/\lambda\right)\left[u_1'\left[\mathbb{E}\left(\cdot\right)\right] a_i\left(1 - T_k'\left[\mathbb{E}\left(a_i r_{\underline{i}}^p z\right)\right]\right) + \frac{1}{2}u_1''\left(a_i\right) a_i^2\right]\mathbb{E}\left(r_{\underline{i}}^p z\right) di + o\left(z\right) \\
&+ \frac{1}{\underline{x}}\beta H v\left(\underline{x}\right) z \int_{\underline{i}}^1 \left(\mathit{\Gamma}_i/\lambda\right) u_1'\left[a_i\left(1 + \mathbb{E}\left(r_i^p z\right)\right) - T_k\left(a_i \mathbb{E}\left(r_i^p z\right)\right) - \kappa\left(x_i - \underline{x}\right) z\right] di + o\left(z\right) \\
&= \frac{1}{\underline{x}}\beta H \zeta_{\underline{i}}^{\mathbb{E}(r^p z),\underline{x}} \int_0^{\underline{i}} \mathbb{E}\left[g_{i,1}\left(\cdot\right)\right] \frac{dlog\left[\mathbb{E}\left(u_1'\left(\cdot\right)\right)\right]}{dlog\left[\mathbb{E}\left(r_{\underline{i}}^p z\right)\right]} di + \frac{1}{\underline{x}}\beta H v\left(\underline{x}\right) z \int_{\underline{i}}^1 g_{i,1}\left[\mathbb{E}\left(\cdot\right)\right] di + o\left(z\right),
\end{aligned}$$

which shows Equation (86). The first-order revenue effect (Equation (87))

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

collects the effects on the capital income of households below $\underline{i}$ and income effects for all households.

UniCredit Foundation
Piazza Gae Aulenti, 3
UniCredit Tower A
20154 Milan
Italy

Giannantonio De Roni – *Secretary General*
e-mail: giannantonio.deroni@unicredit.eu

Annalisa Aleati - *Scientific Director*
e-mail: annalisa.aleati@unicredit.eu

Info at:
www.unicreditfoundation.org

UniCredit Foundation