TI 2020-038/VI
Tinbergen Institute Discussion Paper

Optimal Taxation of Capital Income with Heterogeneous Rates of Return

Aart Gerritsen[1]
Bas Jacobs[1]
Alexandra V. Rusu[2]
Kevin Spiritus[1]

[1] Erasmus School of Economics, Erasmus University Rotterdam
[2] DG Taxation and Customs Union, European Commission

Optimal Taxation of Capital Income with Heterogeneous Rates of Return

Aart Gerritsen, Bas Jacobs, Alexandra V. Rusu and Kevin Spiritus*

June 22, 2020

Abstract

There is increasing empirical evidence that people systematically differ in their rates of return on capital. We derive optimal non-linear taxes on labor and capital income in the presence of such return heterogeneity. We allow for two distinct reasons why returns are heterogeneous: because individuals with higher ability obtain higher returns on their savings, and because wealthier individuals achieve higher returns due to scale effects in wealth management. In both cases, a strictly positive tax on capital income is part of a Pareto-efficient dual income tax structure. We write optimal tax rates on capital income in terms of sufficient statistics and find that they are increasing in the degree of return heterogeneity. Numerical simulations for empirically plausible return heterogeneity suggest that optimal marginal tax rates on capital income are positive, substantial, and increasing in capital income.

JEL: H21, H24

Keywords: Optimal taxation, capital taxation, heterogeneous returns

1 Introduction

Income inequality is rising in most parts of the world, in part due to rising inequality in capital income and wealth (Alvaredo et al., 2018). Piketty (2014)'s *Capital in the Twenty-First Century* brought the question of how governments should tax capital back to the center of the policy debate. Arguments against taxation of capital income date back to Mill (1848) and Pigou (1928), who argued that taxes on capital income amount to taxing labor income twice: first when it is earned, second when it is saved. If all inequality in capital income derives only from inequality in labor income, then it is of no independent concern for the optimal taxation of capital income. Taxes on capital income would redistribute income and distort labor supply in the same way as taxes on labor income, but they would also distort saving decisions. Hence, it would be better not to tax capital income at all (Atkinson and Stiglitz, 1976).[1]

*Gerritsen, Jacobs and Spiritus: Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Rusu: Directorate General Taxation and Customs Union, European Commission, B-1049 Brussels. Gerritsen: E-mail agerritsen@ese.eur.nl, Internet https://sites.google.com/site/aartgerritsen. Jacobs: E-mail bjacobs@ese.eur.nl, Internet http://personal.eur.nl/bjacobs. Rusu: E-mail alexandra.v.rusu@gmail.com, Internet https://alexrusu.com. Spiritus: E-mail spiritus@ese.eur.nl, Internet: http://www.kevinspiritus.eu. We thank Spencer Bastani, Giacomo Corneo, Katherine Cuff, André Decoster, Tilman Klumpp, Tuomas Kosonen, Claus Kreiner, Agnar Sandmo, Floris Zoutman and participants of numerous seminars and conferences for helpful comments and discussions. Bas Jacobs acknowledges support from the Netherlands Institute for Advanced Studies (Duisenberg Fellowship 2018-2019). Kevin Spiritus acknowledges the financial support of the Belgian Federal Science Policy Office (BELSPO) via the BRAIN.be project BR/121/A5/CRESUS.

[1]The result that taxes on capital income are undesirable to redistribute income requires that individuals have identical and weakly separable preferences between consumption at different dates and leisure. Intuitively, conditional on labor

The view that taxes on capital income are not helpful for income redistribution has been highly influential in academic and policy debates (Stiglitz, 2018). However, it critically hinges on the assumption that all individuals obtain the same rate of return on their savings, regardless of their earning abilities or wealth. This assumption has become untenable. A large and growing body of empirical evidence shows that people differ in their returns on savings, and that these returns are systematically related to measures of ability and wealth. Importantly, these return differences are persistent and present even after controlling for risk-taking behavior. This evidence strongly suggests that inequality in capital income does not simply derive from inequality in labor income.

There are roughly three strands of empirical literature that speak to the importance of return heterogeneity. First, the most direct evidence simply documents differences in rates of return across the population. A seminal contribution is Yitzhaki (1987), who studies a subset of US tax returns from 1973 and finds that rates of return increase with income. Piketty (2014) documents that universities with larger endowments are able to generate substantially larger returns on their investments than universities with smaller endowments. Saez and Zucman (2016) find the same for all registered US foundations. More recently, Fagereng et al. (2020) and Bach, Calvet, and Sodini (2020) find convincing evidence of significant return heterogeneity on the basis of administrative data on the populations of taxpayers in, respectively, Norway and Sweden over multiple years. Fagereng et al. (2020) find that differences in rates of return are important, persistent, and attributable to individual-specific factors that cannot be explained by observables, such as differences in the allocation of wealth between risky and safe assets. Moving from the 10th to the 90th percentile of the distribution of *gross* financial wealth (excluding non-financial wealth, such as businesses and housing), the average rate of return increases by 1.6 percentage points. This figure is only slightly lower if they restrict attention to safe assets or if they control for the volatility of the underlying portfolio.

Second, a large literature in finance documents that richer individuals tend to make fewer mistakes in their investments. An abundance of evidence shows that individuals do not optimally diversify their portfolios (e.g., Benartzi and Thaler, 2001; Choi, Laibson, and Madrian, 2005; Calvet, Campbell, and Sodini, 2007; Goetzmann and Kumar, 2008; Von Gaudecker, 2015). Furthermore, individuals consistently fail to optimize their financial portfolio even conditional on risk, for example by exposing themselves to excess interest and fee payments (Barber, Odean, and Zheng, 2005; Agarwal et al., 2009; Choi, Laibson, and Madrian, 2010, 2011). Investment mistakes may also be facilitated by fraudulent financial intermediaries that cater to financially unsophisticated clients (Egan, Matvos, and Seru, 2018). Unsurprisingly, investment mistakes are linked to individuals' financial literacy or sophistication, which itself is positively associated with education and wealth (e.g., Van Rooij, Lusardi, and Alessie, 2011; Lusardi and Mitchell, 2011; Lusardi, Michaud, and Mitchell, 2017). See also Campbell (2016) for a recent overview on mistakes in household finance. A natural implication of this evidence is that richer individuals obtain higher rates of return on their savings.

Third, recent research suggests that return heterogeneity is necessary to reconcile life-cycle models with observed patterns of wealth inequality. In particular, Benhabib, Bisin, and Luo (2019) and Gabaix et al. (2016) argue that return heterogeneity is needed to explain the dynamics of the fat, right tail of the US wealth distribution. Importantly, Gabaix et al. (2016) emphasize both "type dependence" and "scale dependence" in return heterogeneity. That is, they argue that rates of return could depend on both underlying the individual type – e.g., cognitive ability – and the level of individual wealth. Lusardi, Michaud, and Mitchell (2017) and Kacperczyk, Nosal, and Stevens (2018) emphasize the importance of

income, all individuals then save the same amount (Atkinson and Stiglitz, 1976).

heterogeneity in returns due to differences in financial sophistication to explain inequality in wealth and capital income. Indeed, Lusardi, Michaud, and Mitchell (2017) suggest that 30-40 percent of inequality in US retirement wealth can be explained by return heterogeneity.

We analyze the implications of return heterogeneity for optimal non-linear taxes on both labor and capital income. We study a two-period version of the Mirrlees (1971) model. Individuals differ in their ability and choose how much to work and how much to save in the first period of their life cycle. They consume all their savings and the returns on their savings in the second period of their life cycle.[2] Labor income is a function of labor supply and ability, whereas capital income is a reduced-form function of savings and ability, which is able to capture a number of microfoundations for return heterogeneity. The government can only observe labor income and capital income, and not ability. As a result, it must rely on distortionary taxes on labor and capital income to optimally redistribute income. We abstract from risk to focus our analysis solely on the implications of systematic heterogeneity in returns. We derive optimal taxes for two different microfoundations of return heterogeneity. In the first microfoundation, rates of return reflect type dependence as they are determined by ability. In the second microfoundation, rates of return reflect scale dependence and are determined by the amount of wealth. The two microfoundations generate two different reasons to tax capital income.

First, we consider type-dependent returns. Individuals have access to both a freely traded asset with a fixed rate of return, and a closely held asset with decreasing returns to capital and positive marginal returns to ability. Positive returns to ability may reflect a positive association between earning ability and entrepreneurial talent. In equilibrium, individuals equate the marginal rates of return of freely traded and closely held assets. While everyone therefore invests at the same *marginal* rates of return, individuals with higher ability obtain higher *average* rates of return. The resulting capital income increases with ability – for given savings. Capital income therefore reveals information about ability in addition to what is revealed by labor income. This implies that the government should tax both capital income and labor income at positive marginal rates to optimally redistribute income. The optimal tax on capital income trades off (additional) redistributional gains against distortions in savings.

We derive an expression for the Pareto-efficient structure of taxes on capital income and labor income. The optimal dual tax structure equates the marginal distortionary costs of both taxes for the same amount of redistribution. The optimal dual tax structure does not depend on social welfare weights, but only on tax wedges and elasticities of labor income and capital income. The critical term in the optimal tax formula is the elasticity of capital income with respect to ability, conditional on labor income. This elasticity measures the degree of return heterogeneity. In the absence (presence) of return heterogeneity, the elasticity and the optimal tax on capital income are zero (strictly positive). Hence, our model nests Atkinson and Stiglitz (1976) as a special case. More generally, the larger the degree of return heterogeneity, the larger the elasticity of capital income with respect to ability, and thus the larger the optimal tax on capital income.

Second, we consider scale-dependent returns. Individuals with more wealth obtain higher marginal rates of return due to positive scale effects. Implicitly, there is a failure of the capital market that prevents individuals with low wealth ('the poor') from investing in assets with high returns. In particular, it would be mutually beneficial for the poor to lend funds to the rich and receive part of their superior returns, but a market failure prevents these transactions from taking place. We show that a positive

[2]The life-cycle structure of our model is similar to that of Ordover and Phelps (1979), who also analyse optimal taxes on capital income in a two-period OLG version of the Mirrlees (1971) model. We assume that preferences over consumption and leisure are identical and separable, so that optimal taxes on capital income would be zero in the absence of return heterogeneity, see also Atkinson and Stiglitz (1976).

tax on capital income is part of a policy that mimics these missing transactions and thus alleviates the market failure. Specifically, the government could reduce marginal taxes on labor income and raise marginal taxes on capital income. Such a policy would transfer funds from the poor to the rich in the first period and transfer funds back from the rich to the poor in the second period – thus mimicking the missing market transaction. The optimal tax on capital income is positive and trades off the benefits of alleviating the market failure against the costs of distorting savings.

We derive an intuitive ABC-style formula for Pareto-efficient taxes on capital income in the spirit of Diamond (1998) and Saez (2001). The main difference between the usual ABC-formula and ours is that social welfare weights are replaced by marginal returns to savings. Thus, the optimal marginal tax on capital income is inversely related to the elasticity of capital income and the relative hazard rate of the capital income distribution. It is increasing in the difference between the average marginal rate of return for people with a relatively high capital income and the average marginal rate of return for the whole population. The latter implies that optimal taxes on capital income are zero in the absence of return heterogeneity and increasing with the extent of return heterogeneity. Again, our model nests Atkinson and Stiglitz (1976) as a special case. We also derive a simple formula for the optimal top tax rate on capital income.

Finally, we numerically simulate our model to obtain a quantitative sense of the importance of return heterogeneity for optimal tax policy. We mainly focus our numerical simulations on the case of type-dependent returns. We calibrate our model on the basis of US data on the distribution of income, but model return heterogeneity by using Norwegian estimates from Fagereng et al. (2020). We find positive and substantial optimal taxes on capital income in most of our simulations. In our baseline simulation, optimal tax rates on capital income are on average around 14 percent, while the optimal top tax rate at the 99th income percentile is around 30 percent. Moreover, the optimal marginal tax rate on capital income is generally increasing in capital income. In the case of scale-dependent returns, we only calibrate results for the top of the income distribution and find an optimal top tax rate of around 23 percent. However, all of our results are sensitive to both the extent of return heterogeneity and the elasticity of capital income with respect to the marginal rate of return. Because there is a lack of consensus on their empirical magnitudes, we cannot make strong claims on the exact level of the optimal tax rate on capital income. Nevertheless, our numerical simulations demonstrate that empirically plausible return heterogeneity may call for positive and substantial taxes on capital income.

The remainder of the paper is organized as follows. In Section 2, we discuss earlier results on optimal taxation of capital income and indicate how we contribute to this large literature. In Section 3, we introduce and discuss the theoretical setting of our paper. In Section 4, we explicitly show how our model is able to capture two different plausible microfoundations of return heterogeneity. In Section 5, we derive and discuss the optimal non-linear taxes on labor and capital income. Section 6 provides numerical simulations of optimal taxes on labor and capital income under realistic assumptions on return heterogeneity. A final section concludes. The Appendix contains derivations and proofs of all propositions and lemmas.

2 Related literature

A number of key papers identify settings in which optimal taxes on capital income are zero (Atkinson and Stiglitz, 1976; Chamley, 1986; Judd, 1985). Much of the subsequent literature on capital taxation explores motives to tax capital income at positive rates. Taxes on capital income may be optimal to alleviate distortions of labor taxes on labor supply (Corlett and Hague, 1953; Atkinson and Stiglitz, 1976;

Erosa and Gervais, 2002; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018), human capital (Jacobs and Bovenberg, 2010), or to contain tax shifting between labor and capital income (Christiansen and Tuomala, 2008; Reis, 2010). They may be an efficient instrument for redistribution if saving preferences are increasing with ability (Mirrlees, 1976; Saez, 2002; Diamond and Spinnewijn, 2011), if individuals have heterogeneous preferences for wealth itself (Saez and Stantcheva, 2018) or if endowments and inheritances positively correlate with labor income but cannot be directly taxed (Cremer, Pestieau, and Rochet, 2001).[3] In models with overlapping generations, taxes (or subsidies) on capital income may be helpful to correct dynamic inefficiencies in capital accumulation (Ordover and Phelps, 1979; Atkinson and Sandmo, 1980; King, 1980). Finally, there are papers that derive optimally positive taxes on capital income in settings with idiosyncratic shocks to labor productivity and missing capital or insurance markets (Aiyagari, 1995; Hubbard and Judd, 1986; Diamond and Mirrlees, 1978; Golosov, Kocherlakota, and Tsyvinski, 2003; Conesa, Kitao, and Krueger, 2009; Jacobs and Schindler, 2012).[4]

Our paper is most closely related to a small number of papers that also study optimal taxation with heterogeneous returns to capital. Stiglitz (1985, 2000, 2018) has conjectured but not formally shown that optimal taxes on capital income are positive if rates of return depend on ability. We confirm this conjecture. Gahvari and Micheletto (2016) and Kristjánsson (2016) study the two-type optimal tax framework of Stiglitz (1982), and show that optimal taxes on capital income are positive if rates of return are higher for the high-ability type. We contribute to these papers in a number of ways. First, we show that optimal taxes on capital income are also positive if rates of return are an increasing function of wealth itself, i.e., if returns are scale dependent rather than type dependent. The government can then improve the allocation of capital by reducing taxes on labor income in the first period and raising taxes on capital income in the second period. In contrast, Gahvari and Micheletto (2016) and Kristjánsson (2016) conclude that scale dependency of returns does *not* provide a reason to tax capital income. This conclusion is driven by their assumption that all taxes are levied in the same period. Second, we study an economy with a continuum of types as in Mirrlees (1971). This allows us to derive meaningful optimal tax formulas in terms of sufficient statistics, as well as gain more insight into the shape of the optimal non-linear tax schedule on capital income. Third, we derive conditions for the Pareto-efficient structure of taxes on capital income and labor income that does not depend on social welfare weights. Fourth, we provide numerical simulations of optimal non-linear taxes on capital income, and show that they are increasing with capital income.

Guvenen et al. (2019) numerically study optimal linear tax systems in a macroeconomic model with overlapping generations that differ in their returns to capital, while we analytically derive, interpret and simulate optimal non-linear taxes in a two-period life-cycle model. The way they model return heterogeneity is similar to our first microfoundation with common returns to a freely traded asset and type-dependent returns to a closely-held asset. Moreover, they allow for borrowing constraints. In the absence of borrowing constraints, they find that optimal taxes on capital income are positive. This conforms to our findings in the first microfoundation: we both find an optimal dual tax structure that smooths out distortions over the labor and capital tax bases. However, if borrowing constraints are binding, Guvenen et al. (2019) find that individuals are not able to equate the marginal returns on the freely traded and closely-held assets. This leads to a *misallocation* of capital in which constrained individuals obtain a higher rate of return than unconstrained individuals. Comparable to our second

[3] Saez and Stantcheva (2018) argue that taxing capital income could be desirable if individuals derive utility from wealth. They extend their results to the case where returns are heterogeneous. However, it not clear whether return heterogeneity increases or decreases the optimal tax on capital income, or even whether it affects the optimal tax at all.

[4] See surveys of the literature by Diamond and Saez (2011), Jacobs (2013), and Bastani and Waldenström (2020).

microfoundation with scale-dependent returns, the government then wants to use the tax system to redistribute from low- to high-return individuals. In the case of Guvenen et al. (2019), this is done by setting *lower* and even *negative* taxes on capital income, or positive taxes on wealth. Intuitively, constrained individuals with high rates of return are more likely to have a high level of capital income. Thus, the government can redistribute towards high-return individuals by reducing taxes on capital income. In contrast, we find that misallocation of capital provides a reason for *higher* taxes on capital income. Intuitively, the government wants to redistribute from poor low-return to rich high-return individuals before investments take place, and revert this redistribution once returns are realized. This is achieved by reducing marginal taxes on (early-life) labor income and increasing marginal taxes on (later-life) capital income.

Finally, there are papers that focus on optimal taxation if returns differ due to risk and portfolio choice. Varian (1980) shows that optimal taxes on capital income are positive if returns to savings feature idiosyncratic risk, and tax revenues can be returned in lump-sum fashion. Intuitively, taxes on capital income then provide social insurance by redistributing capital income from the lucky to the unlucky. Similarly, Gordon (1985) studies optimal taxation of capital income if individuals invest in both risk-free and risky assets. They find that taxes on capital income yield no insurance gains if there is only aggregate risk in capital returns and revenues are used to finance state-contingent lump-sum transfers. However, Christiansen (1993) and Schindler (2008) show that capital taxes are still optimal if the revenue is used to finance public goods. The optimal tax on capital income then balances the risk in private consumption against the risk in public good provision. Spiritus and Boadway (2017) study optimal taxes on both normal and above-normal returns to capital. Taxing above-normal returns insures idiosyncratic risk in capital income. A tax on the normal rate of return might be desirable if this leads to portfolio reallocation towards assets with idiosyncratic risk. All these studies suggest that the case for a positive optimal tax on capital income would be strengthened if we would allow for risky returns and portfolio choice.

3 Model

3.1 Individual behavior

Individuals are assumed to live for two periods. Individuals differ only in their innate ability $n \in [0, \infty)$, drawn from a cumulative distribution function $F(n)$ with density $f(n)$. Individual ability determines labor productivity and possibly affects returns to savings. As it is the only source of heterogeneity, we denote individuals by their ability n. In the first period, individual n supply labor l^n and earns labor income $z^n \equiv nl^n$. First-period income is spent on taxes on labor income T^n, consumption c_1^n, and savings a^n. Thus, we can write first-period consumption as:

$$c_1^n = z^n - T^n - a^n. \tag{1}$$

Savings yield deterministic capital income y^n, which depends on the amount of savings and, potentially, on individual ability: $y^n = y(a^n, n)$. As we show later, this formulation allows us to capture plausible microfoundations of return heterogeneity related to closely-held businesses and scale economies in wealth investment. The case where returns on the assets from a closely-held business are increasing in the owner's ability could be captured by $y_n > 0$. Returns that are increasing in total wealth of an individual could be captured by $y_{aa} > 0$. In the latter case, individuals generally differ in their marginal rate of return y_a. Thus, we implicitly allow for capital-market failures, such that differences in marginal rates of return are not necessarily arbitraged away. Taxes on capital income are denoted by τ^n, and second-period consumption equals the sum of savings and after-tax capital income:

$$c_2^n = a^n + y(a^n, n) - \tau^n. \tag{2}$$

T^n is a non-linear tax function of labor income z^n, and τ^n is a non-linear tax function of capital income y^n. We parameterize the tax schedules in a way that allows us to study the effects of exogenous shifts in their slopes and intercepts. This later helps us define behavioral elasticities and social welfare weights.[5] We write the tax schedules as the following functions:

$$T^n = T(z^n, \rho^T, \sigma^T) \equiv \tilde{T}(z^n) + \rho^T + \sigma^T z^n, \tag{3}$$

$$\tau^n = \tau(y^n, \rho^\tau, \sigma^\tau) \equiv \tilde{\tau}(y^n) + \rho^\tau + \sigma^\tau y^n, \tag{4}$$

where ρ^T and ρ^τ are parameters that shift the intercepts of the tax schedules, and σ^T and σ^τ are parameters that shift the slopes of the tax schedules. The parameterization does not impose any restrictions on the tax schedules because $\tilde{T}(z^n)$ and $\tilde{\tau}(y^n)$ are fully non-linear functions of the tax base.

Individuals derive utility from first- and second-period consumption, and disutility from labor supply. The utility function of individual n can be written as:

$$U^n = u(c_1^n, c_2^n) - v(z^n/n). \tag{5}$$

Utility of consumption $u(\cdot)$ is increasing, concave, and twice continuously differentiable. Disutility of work $v(\cdot)$ is increasing, strictly convex and twice continuously differentiable. Utility is separable between consumption and labor supply, so there is no reason to tax capital income in the absence of return heterogeneity (Atkinson and Stiglitz, 1976). Substituting first- and second-period consumption and the parameterized tax schedules into the utility function and optimizing over savings and labor income yields the following first-order conditions:

$$\frac{v'(z^n/n)}{u_1(c_1^n, c_2^n)} = (1 - T'(z^n, \rho^T, \sigma^T))n, \tag{6}$$

$$\frac{u_2(c_1^n, c_2^n)}{u_1(c_1^n, c_2^n)} = \frac{1}{1 + (1 - \tau'(y(a^n, n), \rho^\tau, \sigma^\tau))y_a(a^n, n)} \equiv \frac{1}{R^n}. \tag{7}$$

We denote marginal tax rates by a prime, so that $T'(z^n, \rho^T, \sigma^T) \equiv \partial T(z^n, \rho^T, \sigma^T)/\partial z^n$ and $\tau'(y^n, \rho^\tau, \sigma^\tau) \equiv \partial \tau(y^n, \rho^\tau, \sigma^\tau)/\partial y^n$. Other partial derivatives are denoted by a subscript. Thus, $u_1(\cdot)$ and $u_2(\cdot)$ are the marginal utility of first- and second-period consumption, and $y_a(\cdot)$ denotes the marginal rate of return. Eq. (6) shows that the marginal rate of substitution between first-period consumption and leisure must equal the marginal after-tax wage rate. Eq. (7) shows that the marginal rate of substitution between first- and second-period consumption must equal the individual's discount factor. We define the inverse

[5] It is not uncommon to parameterize non-linear tax schedules to derive the comparative statics, see, e.g., Christiansen (1981), Immervoll et al. (2007), Jacquet, Lehmann, and Van der Linden (2013), Gerritsen (2016).

of the discount factor – or one plus the after-tax rate of return – as $R^n \equiv 1 + (1-\tau')y_a$.[6]

We impose a number of assumptions that help us derive the optimal non-linear tax schedules. First, we require that both tax schedules are twice continuously differentiable. This ensures that the individual first-order conditions are differentiable. Second, we assume that second-order conditions are satisfied, and that eqs. (6) and (7) describe a unique and global maximum for utility. This guarantees that individual behavior is differentiable and thus that marginal changes in taxes lead to marginal responses in earnings. It also implies that the equilibrium values of both tax bases y^n and z^n are monotonically increasing in ability n. These assumptions correspond to Assumption 2 in Jacquet and Lehmann (2017).

3.2 Behavioral elasticities

Behavioral elasticities of the tax bases play an important role in the optimal tax expressions that we derive below. To define these elasticities, we first write the tax bases as functions of the tax parameters. The first-order condition for savings in eq. (7), together with the definitions of first- and second-period consumption (1) and (2), implicitly determines equilibrium savings as a function of labor income, tax parameters, and ability. This allows us to write equilibrium savings as $a^n = \tilde{a}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$, where the superscript c indicates conditionality on labor income z^n. Since capital income is a function of savings and ability, we can write equilibrium capital income as a function of the same arguments $y^n = \tilde{y}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n) = y(\tilde{a}^c, n)$. *Both* first-order conditions in eqs. (6) and (7), together with the definitions of first- and second-period consumption, determine labor income as a function of tax parameters and ability. This allows us to write equilibrium labor income as $z^n = \tilde{z}(\rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$.[7]

We define the *compensated* elasticity of labor income with respect to the net-of-tax rate for each individual n as:

$$e_z^n \equiv -\left(\frac{\partial \tilde{z}}{\partial \sigma^T} - z^n \frac{\partial \tilde{z}}{\partial \rho^T}\right)\frac{1-T'}{z^n}. \tag{8}$$

The elasticity in eq. (8) measures the percentage change in labor income if the net-of-tax rate $1-T'$ is exogenously raised by one percent, while utility is kept constant. It captures the total impact on labor income, taking into account the effect a change in earnings can have on the marginal tax rate if the tax function is non-linear.[8] The term within brackets gives the Slutsky decomposition of the compensated response in labor income to an increase in marginal taxes.[9]

We define the compensated elasticity of capital income with respect to the after-tax rate of return for each individual n as:

$$e_{y|z}^n \equiv -\frac{1}{y_a}\left(\frac{\partial \tilde{y}^c}{\partial \sigma^\tau} - y^n \frac{\partial \tilde{y}^c}{\partial \rho^\tau}\right)\frac{R^n}{y^n}. \tag{9}$$

The elasticity in eq. (9) measures the percentage change in capital income if the after-tax rate of return $R^n = 1 + (1-\tau')y_a$ is exogenously raised by one percent, while utility is kept constant. It captures the total impact on capital income, taking into account the effect a change in capital income can have on

[6] In what follows, we suppress function arguments for brevity unless this is likely to cause confusion.

[7] We denote equilibrium functions for the tax bases with a tilde. We do this to distinguish equilibrium capital income $\tilde{y}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$ from capital income as a function of savings and ability $y(a^n, n)$.

[8] In terms of Jacquet and Lehmann (2017), e_z^n is a 'total elasticity' rather than a 'direct elasticity.' The elasticity measures the effect on labor income of a given change in the tax parameters σ^T and ρ^T rather than a given change in the marginal tax rate $T'(z, \rho^T, \sigma^T)$. Total elasticities are also used by, e.g., Jacquet, Lehmann, and Van der Linden (2013), Jacobs and Boadway (2014), Gerritsen (2016), and Scheuer and Werning (2017).

[9] We formally prove this in Appendix C.1.

the marginal tax rate if the tax function is non-linear. Again, the term within brackets is the Slutsky decomposition of the compensated response of capital income to an increase in the marginal tax rate. Furthermore, $e^n_{y|z}$ is a *conditional* elasticity, in that it measures the behavioral change in capital income while holding labor income constant.

Finally, we define the elasticities of labor and capital income with respect to ability as:

$$\xi^n_z \equiv \frac{\partial \tilde{z}}{\partial n}\frac{n}{z^n}, \qquad \xi^n_{y|z} \equiv \frac{\partial \tilde{y}^c}{\partial n}\frac{n}{y^n}. \tag{10}$$

The first elasticity ξ^n_z measures the percentage change in labor income due to a one percent increase in ability. The second elasticity $\xi^n_{y|z}$ measures the percentage change in capital income due to a one percent increase in ability, *while holding labor income constant.*

4 Two microfoundations of return heterogeneity

4.1 Type-dependent returns: entrepreneurial investments

It is instructive to consider two plausible microfoundations for capital income $y(a^n, n)$ that could generate heterogeneity in rates of return. These two different microfoundations loosely correspond to what Gabaix et al. (2016) call type-dependent and scale-dependent returns. We first consider type-dependent returns. In particular, we consider an economy in which individuals can invest in two different types of assets. They can invest in a closely held asset that is specific to their type and could be interpreted as entrepreneurial investment. And they can invest in an asset that is freely traded in capital markets.

Individual n invests b^n in the closely held asset. This yields a total return that is a function of invested capital and ability: $\pi^n = \pi(b^n, n)$. The closely held asset exhibits decreasing returns to capital ($\pi_b > 0$ and $\pi_{bb} < 0$) and positive returns to ability ($\pi_n > 0$). The latter assumption reflects the idea that high ability helps to find and select successful business ventures. The remainder of individual savings, $a^n - b^n$, is invested in the freely traded asset, which yields a common, constant rate of return r.

Capital income is now given by:

$$y^n = r(a^n - b^n) + \pi(b^n, n). \tag{11}$$

Individuals allocate their savings over the two assets in a way that maximizes their capital income (provided that the marginal tax on capital income is below 100 percent, $\tau' < 1$). Maximizing y^n in eq. (11) with respect to b^n yields $\pi_b(b^n, n) = r$. Thus, individuals invest in the closely-held asset up to the point at which its marginal return equals that on the commonly traded asset. This implicitly determines entrepreneurial investment as a function of ability alone: $b^n = b(n)$. Substituting this into the equation for capital income (11) yields:

$$y^n = y(a^n, n) = ra^n + \pi(b(n), n) - rb(n). \tag{12}$$

Hence, the general formulation $y^n = y(a^n, n)$ can capture the special case of entrepreneurial investments. Under this microfoundation, capital income is linear in savings and increasing in ability: $y_a = r$ and $y_n > 0$.[10]

[10] Both follow from the partial derivatives of $y(a, n)$ in eq. (12). $y_a = r$ follows trivially. Application of the envelope theorem yields $y_n = \pi_n \geq 0$.

4.2 Scale-dependent returns: scale economies in wealth investment

The second microfoundation of capital income $y(a^n, n)$ relies on scale economies in wealth investment. Scale economies may originate from the fixed costs associated with raising rates of return. For example, an individual needs a savings account with a bank to earn any interest on savings at all. Because banks typically charge their account holders fixed periodic fees, it only makes sense to open an account and obtain a positive rate of return if savings are large enough to cover these fixed fees. Moreover, to participate in higher-yielding assets such as equity, one needs to invest in at least some basic financial knowledge or acquire the costly services of a wealth manager. Again, it only makes sense to pay for these higher yields if the invested wealth is sufficiently large. As a consequence, individuals with more wealth are likely to obtain higher rates of return.

We can capture scale effects in our model by assuming that individuals invest x^n of their savings to raise the returns on the remainder of their savings. These investments consist of search costs, fixed fees, and the costs of obtaining financial know-how. This leaves an amount $a^n - x^n$ to be invested at a rate of return $r(x^n) \geq 0$ with $r'(x^n) \geq 0$. We assume that all investment costs are deductible from the capital tax base but not separately observed by the government and thus not separately taxed.[11] Taxable capital income is then given by:

$$y^n = r(x^n)(a^n - x^n) - x^n. \tag{13}$$

Individuals invest in financial services to maximize their capital income. Maximizing y^n in eq. (13) with respect to x^n yields $r'(x^n)(a^n - x^n) = 1 + r(x^n)$.[12] The left-hand side gives the gains from investing one more unit of resources in obtaining a higher rate of return. The right-hand side denotes the opportunity costs of doing so. The equilibrium condition implicitly determines investment costs as a function of savings $x^n = x(a^n)$, with $x'(a^n) \geq 0$. Intuitively, the larger one's wealth, the stronger are the incentives to increase the rate of return. Substituting this into the expression for capital income yields:

$$y^n = y(a^n, n) = r(x(a^n))(a^n - x(a^n)) - x(a^n). \tag{14}$$

Hence, the general formulation $y^n = y(a^n, n)$ also captures scale economies in wealth investment. In that case, capital income is convex in savings and does not (directly) depend on ability: $y_a \geq 0$, $y_{aa} \geq 0$, and $y_n = 0$.[13]

Individuals with different levels of wealth face different marginal rates of return and therefore different *marginal rates of transformation* between first- and second-period consumption. The costs x can be interpreted as the costs of entering a specific financial market in which assets yield a rate of return $r(x)$. Thus, individuals with different levels of wealth effectively invest in segmented financial markets. This means that there are potential Pareto-improving trades in the capital market that do not materialize. To see this, imagine that a rich high-return individual borrows funds from a relatively poor low-return individual at some intermediate interest rate. Such a loan would be mutually beneficial because the poor individual obtains a higher return, while the rich individual pockets the difference

[11] Investment funds typically subtract their fees from the payout to the participants. This effectively makes the investment fees tax deductible for the owner of the wealth.

[12] The second-order condition that ensures an interior solution is given by $r''(x^n)(a^n - x^n) < 2r'(x^n)$. It is intuitively plausible that there is an upper limit to which the rate of return can rise by investing more and more in search costs, wealth management fees, and financial know-how. If the rate of return $r(x)$ is indeed increasing in x at a decreasing rate, then the second-order condition is always satisfied.

[13] This follows from the partial derivatives of eq. (14). $y_n = 0$ follows trivially. Application of the envelope theorem yields $y_a = r(x(a^n)) \geq 0$ and, hence, $y_{aa} = r'x' \geq 0$.

between the rate of return and the interest rate charged to the poor individual. Thus, implicit in the microfoundation is a market failure that keeps relatively poor individuals from accessing the higher-yielding investment opportunities of the rich.

5 Optimal taxation

5.1 Instrument set

We assume that the government can observe labor income and capital income at the individual level. Hence, the government can implement non-linear taxes on labor income and capital income. We also assume that capital income and labor income are taxed separately. In doing so we follow most of the literature (e.g., Saez and Stantcheva, 2018). As a result, the marginal tax rate on one tax base does not depend on the size of the other tax base.[14]

We rule out taxes on consumption and wealth as this would enable the government to tax away all heterogeneous returns with zero distortions.[15] In practice, 100 percent taxation of excess returns via wealth or consumption taxes will surely result in tax avoidance and evasion due to cross-border shopping, international mobility of capital, and reduced entrepreneurial efforts. Hence, our model realistically captures the main policy trade-off for the optimal taxation of capital income with heterogeneous returns, while avoiding complexities with modeling cross-border shopping, capital mobility or entrepreneurial effort.

5.2 Social welfare and government budget constraints

The government sets and fully commits to taxes on labor and capital income. Social welfare is an additive, concave function of individual utilities:

$$\mathcal{W} = \int_0^\infty W(U^n) f(n)\mathrm{d}n, \quad W'(U^n) > 0 \quad W''(U^n) \leq 0. \tag{15}$$

Social preferences for income redistribution are captured by concavity of either the welfare function W or the utility function U^n.

The government levies taxes on labor income in the first period, and taxes on capital income in the second period. We consider the net asset position of the government as exogenously fixed. Thus, the government cannot shift the tax burden from one period to the other by issuing new (or repurchasing old) bonds. As a result, the government faces binding budget constraints in both the first and the second period:

$$\mathcal{B}_1 = \int_0^\infty T(z^n, \rho^T, \sigma^T) f(n)\mathrm{d}n - g_1 = 0, \tag{16}$$

$$\mathcal{B}_2 = \int_0^\infty \tau(y^n, \rho^\tau, \sigma^\tau) f(n)\mathrm{d}n - g_2 = 0, \tag{17}$$

[14] Renes and Zoutman (2014) show that separable tax schedules are sufficient to implement the full second-best optimum if there are no market failures, the government has a welfarist objective, and the population is characterized by one-dimensional heterogeneity. These assumptions are fulfilled in our first microfoundation with type-dependent returns. However, these assumptions are no longer fulfilled in our second microfoundation with scale-dependent returns due to the implicit failure of the capital market. Hence, separate tax schedules on labor and capital income may not generally implement the full second-best optimum.

[15] This can best be seen in the context of type-dependent returns. A 100 percent tax on capital income combined with a subsidy on wealth would tax away excess returns, as would letting consumption taxes and labor subsidies go to infinity.

where g_1 and g_2 are exogenous revenue requirements in periods 1 and 2.

Assuming fixed government assets is innocuous in the first microfoundation, because government debt would be neutral if the government were able borrow and lend at the same rate as all individuals.[16] Government assets are assumed to be fixed primarily in order to prevent infinite borrowing under the second microfoundation, where the government would take over all investment in the entire economy to generate maximum scale effects. Apart from running into inessential but complicated technical digressions to deal with corner solutions, fixing the government asset position is a simple shortcut to model the possible efficiency losses or political-economy distortions associated with managing extremely large public wealth funds.

We denote the shadow prices of first- and second-period government revenue by λ_1 and λ_2, so that the social planner's objective function can be written as:

$$\mathcal{L} = \frac{1}{\lambda_1}\mathcal{W} + \mathcal{B}_1 + \frac{1}{\lambda_1/\lambda_2}\mathcal{B}_2. \tag{18}$$

The government discounts future tax revenue at a rate λ_1/λ_2.

5.3 Excess burdens and social welfare weights

The optimal tax structure depends on the excess burdens and distributional benefits of taxation. We define the marginal excess burden as the revenue loss caused by a *compensated* increase in a marginal tax rate. The marginal excess burdens of taxes on labor and capital income for individual n are given by:

$$E_T^n \equiv -T'\left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}\right) - \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}\right), \tag{19}$$

$$E_\tau^n \equiv -T'\left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau}\right) - \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau}\right). \tag{20}$$

An increase in marginal taxes potentially affects both tax bases, thereby affecting both first- and second-period revenue. Eq. (19) gives the marginal excess burden of the tax on labor income. The first term equals the revenue loss from a compensated response in labor income, and the second term equals the revenue loss from a compensated response in capital income. Eq. (20) gives the marginal excess burden of the tax on capital income. Again, the equation gives the revenue losses from compensated responses in both labor and capital income.

The distributional benefits of taxation can be expressed by means of social welfare weights. We denote the first- and second-period social welfare weights of individual n by α_1^n and α_2^n:

$$\alpha_1^n \equiv \frac{W'(U^n)u_1}{\lambda_1} - T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} - \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}, \tag{21}$$

$$\alpha_2^n \equiv \frac{W'(U^n)u_2}{\lambda_1} - T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau} - \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau}. \tag{22}$$

[16] Ricardian equivalence applies even though taxes are distortionary, since the government has access to a non-distortionary marginal source of public finance in each period. Hence, the government does not need to introduce tax distortions to steer the intertemporal allocation. See also Werning (2007).

The social welfare weights consist of the (monetized) welfare gains of providing individual n with an additional unit of income in period 1 or 2, and the change in revenue due to the income effects on both tax bases.

5.4 Optimal tax schedules

We solve for the optimal non-linear taxes on labor and capital income by using the Euler-Lagrange formalism, which is a mathematically rigorous version of the heuristic tax-perturbation approach pioneered by Saez (2001), and recently extended and amended by Golosov, Tsyvinski, and Werquin (2014), Gerritsen (2016), and Lehmann et al. (2018). In particular, the Euler-Lagrange formalism employs the calculus of variations to analyze the welfare effects of small perturbations in the tax schedules on labor and capital income. In the optimum, such tax perturbations should have no effect on social welfare.

We denote the density of labor income by $h(z)$ and the density of capital income by $g(y)$. The following Lemma presents optimality conditions for marginal taxes on labor and capital income.

Lemma 1. *In the tax optimum, the following two conditions characterize the optimal marginal tax rates on labor and capital income for all income levels z^n and y^n:*

$$E_T^n h(z^n) = \int_{z^n}^{\infty} (1 - \alpha_1^m) h(z^m) \mathrm{d}z^m, \tag{23}$$

$$E_\tau^n g(y^n) = \int_{y^n}^{\infty} \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^m \right) g(y^m) \mathrm{d}y^m. \tag{24}$$

Proof. See Appendix A. □

The conditions in Lemma 1 are intuitively straightforward. The left-hand sides of these equations represent distortions of labor supply and savings, while the right-hand sides represent mechanical revenue gains, utility losses, and income effects on the tax bases. Consider, first, a small change to the marginal tax rate on labor income in a small interval around income z^n. This tax change distorts labor supply for all individuals with income around z^n. The excess burden associated with this distortion is given by the left-hand side of eq. (23). The perturbation also raises tax revenue from individuals who earn more than z^n. The redistributional gains of this are given by the right-hand side of eq. (23). Analogously, the left-hand side of eq. (24) gives the marginal excess burden of raising the marginal tax rate on capital income around y^n, and the right-hand side gives its redistributional gains. In the optimum, the marginal distortionary costs of raising a marginal tax rate on either labor or capital income should thus be equal to its marginal redistributional gains.

5.5 Optimal taxation of labor income

Lemma 1 expresses the optimal tax schedules in terms of marginal excess burdens and redistributional gains of taxation. To gain more insight into the shape of the optimal tax schedules, we write them in terms of wedges, elasticities, the income distribution, and social welfare weights. The following Proposition establishes the optimal tax wedge on labor income. For notational convenience, we suppress the tax parameters from the function arguments of the tax schedules, so that marginal tax rates at income levels z and y are written as $T'(z)$ and $\tau'(y)$. Moreover, we suppress the superscripts n in view of the perfect mapping between ability and labor and capital income.

Proposition 1. *The optimal tax wedge on labor income for all levels of labor income z is given by:*

$$\frac{T'(z)}{1-T'(z)} + \frac{sy_a\tau'(y)}{\lambda_1/\lambda_2} = \frac{1}{e_z}\frac{1-H(z)}{zh(z)}\left(1-\bar{\alpha}_1^+(z)\right), \tag{25}$$

where $s \equiv (\partial\tilde{a}^c/\partial z)/(1-T'(z))$ is the marginal propensity to save out of net income, $H(z)$ is the cumulative distribution function of labor income, and $\bar{\alpha}_1^+(z) \equiv \int_z^\infty \alpha_1 h(z^*)\mathrm{d}z^*/(1-H(z))$ is the average first-period social welfare weight of individuals that earn more than z.

Proof. See Appendix B. □

The left-hand side of eq. (25) gives the tax wedge on labor income for an individual with income z. To see this, consider a unit increase in after-tax labor income. This implies a $1/(1-T')$ increase in gross labor income, which leads to a revenue gain of $T'/(1-T')$. Moreover, it raises savings by s and capital income by $y_a s$, yielding a second-period revenue gain of $y_a s\tau'$, which the government discounts at a rate λ_1/λ_2. The right-hand side of eq. (25) is the standard expression for the optimal tax wedge on labor, see also Mirrlees (1971), Diamond (1998), and Saez (2001). The optimal tax wedge on labor income is decreasing in the elasticity of labor income at z, e_z, the relative hazard rate of the income distribution at z, $zh(z)/(1-H(z))$, and the average of the social-welfare weights of individuals who earn more than z, $\bar{\alpha}_1^+(z)$. The only material difference with, e.g., Saez (2001), is that the tax wedge on labor income contains not only the tax on labor income, but also the tax on capital income. This is because a reduction in labor income causes individuals to save less, thereby lowering revenue from taxes on both labor and capital income. If the marginal propensity to save is zero ($s=0$), reductions in labor income do not reduce future consumption, so that the standard Saez-formula results.

5.6 Optimal taxation of capital income

In this subsection, we present and discuss our main theoretical results: the expressions for the optimal tax on capital income in the presence of return heterogeneity. We first discuss the case of type-dependent returns, in which return heterogeneity originates from closely held assets. We then discuss the case of scale-dependent returns, in which return heterogeneity originates from scale economies in wealth investment. We end with a more general formulation of the optimal tax on capital income that captures both microfoundations as special cases.

5.6.1 Type-dependent returns ($y_a = r$, $y_n \geq 0$)

The following proposition establishes the optimal marginal tax rate on capital income if returns are type dependent.

Proposition 2. *If capital income is linear in savings but increasing in ability ($y_a = r$ and $y_n \geq 0$ for all individuals), then the optimal marginal tax rate on capital income for every level of capital income y is implied by:*

$$\left(\frac{y_a\tau'(y)}{1+y_a}\right)e_{y|z} = \left(\frac{T'(z)}{1-T'(z)} + \frac{sy_a\tau'(y)}{1+y_a}\right)e_z\left(\frac{\xi_{y|z}}{\xi_z}\right) \geq 0. \tag{26}$$

The inequality is strict only if $\xi_{y|z} > 0$, which holds if and only if $y_n(a,n) > 0$.

Proof. See Appendix C. □

Lemma 1 showed that the marginal tax rate on capital income at income level y has five welfare-relevant effects. For individuals with income above y, it 1) reduces utility, 2) leads to income effects on the tax bases, and 3) mechanically raises government revenue. For individuals who earn capital income y, it 4) distorts labor supply and 5) distorts savings. These marginal costs and benefits cancel out in the tax optimum. A marginal tax on labor income that causes the same utility losses also yields the same income effects and mechanical revenue gains. Effects 1), 2) and 3) are then equal for a tax on capital income and a tax on labor income. Moreover, a tax on labor income would distort labor supply, but leave savings undistorted. Thus, a positive tax on capital income is only desirable if – for the same utility losses – it distorts labor supply less than a tax on labor income. The optimal tax on capital income then equates the marginal costs of distorting savings to the marginal benefits of distorting labor supply *less* than a tax on labor income. This is shown in eq. (26).

The left-hand side of eq. (26) gives the deadweight loss of the savings distortion of a tax on capital income. It equals the tax wedge on savings, $y_a\tau'(y)/(1+y_a)$, multiplied by the conditional elasticity of capital income with respect to the after-tax interest rate $e_{y|z}$. The right-hand side gives the deadweight loss of the labor supply distortions of a tax on labor income *relative to a distributionally equivalent tax on capital income*. The two terms in brackets represent the tax wedge on labor income. It is multiplied by e_z to form the marginal deadweight loss of a tax on labor income. This is multiplied by the ratio of ability elasticities $\xi_{y|z}/\xi_z$, which captures the degree to which the tax on capital income distorts labor supply *less* than the tax on labor income for the same amount of income redistribution. Eq. (26) shows that the optimal tax on capital income is positive only if the ratio $\xi_{y|z}/\xi_z$ is positive. This is the case if and only if capital income is increasing in ability for given savings, such that $y_n(a,n) > 0$. Thus, capital income should be taxed in the presence of type-dependent returns.

Intuitively, the ratio $\xi_{y|z}/\xi_z$ captures the extent to which ability correlates more strongly with capital income than with labor income. Taxes on labor and capital income are second-best, because the government cannot observe and therefore cannot tax ability. Instead, the government wants to tax those tax bases that provide information on ability. If $\xi_{y|z}/\xi_z = 0$, then capital income provides the same information about ability as labor income. In that case, taxes on capital income generate the same distortions on labor supply for the same redistribution in income, and, in addition, distort savings decisions. A tax on capital income is then less desirable than a tax on labor income to redistribute income and should optimally be set to zero. Hence, our model nests Atkinson and Stiglitz (1976) as the special case where returns are equal for all individuals.

If $\xi_{y|z}/\xi_z > 0$, capital income reveals more information about ability than labor income. Starting from a situation without taxes on capital income, but with positive taxes on labor income, introducing a capital tax generates fewer distortions than the tax on labor income for the same redistribution of income. The capital tax generates only a second-order welfare loss in saving, but allows for a first-order welfare gain in labor supply. As a result, in the full optimum, it is desirable to set a positive tax on capital income even if this distorts savings decisions.

The optimality condition in eq. (26) shows that the optimal marginal tax rate on capital income depends on a limited number of key statistics. First of all, it is increasing in the ratio of ability elasticities $\xi_{y|z}/\xi_z$. Second, the optimal tax rate on capital income is decreasing in the conditional elasticity of capital income with respect to the after-tax interest rate $e_{y|z}$. The larger this elasticity, the larger the savings distortions associated with marginal tax rates on capital income. Third, the optimal tax on capital income is increasing in the compensated elasticity of labor income with respect to the net-of-tax rate e_z. The larger this elasticity, the larger are labor supply distortions relative to savings distortions, and thus the more desirable is the tax on capital income relative to a tax on labor income. Fourth, and for the same reason, the optimal tax on capital income is increasing in the tax wedge on labor income. Thus, provided that elasticities are relatively constant over the income distribution, the shape of the optimal tax schedule on capital income tracks the shape of the optimal tax schedule on labor income.

Proposition 2 is closely related to a number of earlier contributions to the literature on the optimal taxation of commodities and of capital income. While Atkinson and Stiglitz (1976) have shown that the government should not use taxes on saving if preferences are homogeneous and separable between consumption and leisure, many subsequent studies have focused on the implications of non-separability and heterogeneity of preferences. Mirrlees (1976) notes that "commodity taxes should bear more heavily on the commodities *high-n* individuals have relatively strongest tastes for" – meaning that we should focus on "the way in which demands change *for given income and labor supply* when n changes." This finding is echoed in subsequent studies by Christiansen (1984), Saez (2002), Diamond and Spinnewijn (2011), and Jacobs and Boadway (2014).[17] Similar to these studies, we also find that capital income should be taxed if it is increasing in ability for given labor income. Contrary to these earlier studies, we show that this argument does not rely on taste heterogeneity or non-separability in preferences. Instead, savings may increase in ability because of empirically plausible heterogeneity in rates of return. This implies that budget constraints rather than saving preferences depend on n for given labor income. In that respect, our findings are also closely related to Cremer, Pestieau, and Rochet (2001), who find that taxes on capital income are desirable if endowments – which are part of the budget constraint – are increasing with ability.

5.6.2 Scale-dependent returns ($y_{aa} \geq 0$, $y_n = 0$)

The following Proposition establishes the optimal marginal tax rate on capital income if rates of return are scale dependent.

Proposition 3. *If capital income is convex in savings, but not directly affected by ability ($y_{aa} \geq 0$ and $y_n = 0$ for all individuals), then the optimal marginal tax rate on capital income for every level of capital income y is implied by:*

$$\frac{\tau'(y)y_a}{1+\bar{y}_a} = \frac{1}{e_{y|z}} \frac{1-G(y)}{yg(y)} \frac{\bar{y}_a^+(y) - \bar{y}_a}{1+\bar{y}_a} \geq 0, \tag{27}$$

where $G(y)$ is the cumulative distribution function of capital income, $\bar{y}_a^+(y) \equiv \int_y^\infty y_a^* g(y^*) \mathrm{d}y^* / (1-G(y))$ is the average marginal rate of return for individuals whose capital income is higher than y, and $\bar{y}_a \equiv \bar{y}_a^+(0) = \int_0^\infty y_a^* g(y^*) \mathrm{d}y^*$ is the average marginal rate of return for all individuals.

Proof. See Appendix C. □

Recall that a marginal tax on capital income at income level y has five welfare-relevant effects. For

[17] Saez (2002) shows that a commodity should be taxed if the consumption-income gradient is steeper over the cross-section of individuals than for any given individual. This is another way of saying that consumption should be increasing in ability for given labor income. Indeed, we could rewrite the ratio of ability elasticities as:

$$\frac{\xi_{y|z}}{\xi_z} = \frac{\partial \tilde{y}^c / \partial n}{\mathrm{d}z/\mathrm{d}n} \frac{z}{y} = \left(\frac{\mathrm{d}y/\mathrm{d}n}{\mathrm{d}z/\mathrm{d}n} - \frac{\partial \tilde{y}^c}{\partial z} \right) \frac{z}{y}.$$

The term within bracket gives the difference between the capital income–labor income gradient over the cross-section of individuals and the same gradient for a given individual.

individuals with income above y, it 1) reduces utility, 2) leads to income effects on the tax bases, and 3) mechanically raises government revenue. For individuals who earn capital income y, it 4) distorts labor supply and 5) distorts savings. These marginal costs and benefits cancel out in the tax optimum.

In contrast to the first microfoundation, capital income does not provide any more information about ability than labor income. This is because capital income is a function of savings only and does not directly depend on ability ($y_n(a^n, n) = 0$). As a result, for the same utility losses, a tax on capital income generates the same labor supply distortion as a tax on labor income. This implies that welfare effects 1), 2), and 4) are equal for a tax on capital income and a tax on labor income.

However, for a given utility loss, the present value of the mechanical revenue gains of a tax on capital income is larger than the value of the mechanical revenue gains of a tax on labor income. To see this, first notice that the discount rate of the government equals the average marginal rate of return $\bar{y}_a$.[18] Intuitively, the government is indifferent between distributing one unit of lump-sum transfers in the first period or $1 + \bar{y}_a$ units of lump-sum transfers in the second period. Second, for a given utility loss, the government can tax away $1 + \bar{y}_a^+$ of mechanical revenue with a tax on capital income for every unit of taxes raised with a tax on labor income. Hence, the present value of the mechanical revenue gains of the tax on capital income exceeds that of the tax on labor income by a factor $(1 + \bar{y}_a^+)/(1 + \bar{y}_a) - 1 = (\bar{y}_a^+ - \bar{y}_a)/(1 + \bar{y}_a)$.

The optimal tax on capital income then equates the marginal costs of distorting savings to the marginal benefits of increasing government revenue by more than a tax on labor income. This is shown in eq. (27). The left-hand side is the tax wedge on savings for individuals with capital income y. Multiplied by $e_{y|z} y g(y)$, it gives the marginal distortionary costs of raising the marginal tax rate at y. On the right-hand side, $1 - G(y)$ gives the share of people that need to pay the marginal tax rate at y. Multiplied by the factor $(\bar{y}_a^+ - \bar{y}_a)/(1 + \bar{y}_a)$, it gives the *additional* mechanical tax revenue from the capital tax as compared to the labor tax.

The optimal tax formula resembles the standard ABC-formula of optimal taxes on labor income (see Proposition 1, as well as Diamond, 1998; Saez, 2001). However, instead of the difference in average welfare weights $(1 - \bar{\alpha}_1^+)$, there is the discounted difference in average marginal rates of return $((\bar{y}_a^+ - \bar{y}_a)/(1 + \bar{y}_a))$. The larger the conditional elasticity of capital income, $e_{y|z}$, the larger the savings distortions and thus the lower the optimal tax on capital income. The larger the inequality in rates of return, $(\bar{y}_a^+(y) - \bar{y}_a)/(1 + \bar{y}_a)$, the larger the marginal benefits of taxing capital income rather than labor income, and thus the higher the optimal tax on capital income. The inverse relative hazard rate of the distribution of capital income, $(1 - G(y))/(yg(y))$, gives weights to the gains of raising revenue through the tax on capital income (numerator) and the cost of distorting savings (denominator). If there are more individuals above y, $(1 - G(y))$ is larger, and the capital tax yields more mechanical revenue. If the concentration of capital income at y is larger, $yg(y)$ is larger, and hence the tax on capital income generates larger distortions.

There are two alternative but equivalent ways of interpreting the role of the capital tax in this framework. First, it helps alleviate a market failure. It would be mutually beneficial if the relatively poor could lend to the rich in the first period and receive part of the superior returns of the rich in the second period. Nevertheless, incomplete capital markets prevent these Pareto-improving transactions from taking place. Such transactions could be *mimicked* by the government's tax policy. In particular, by reducing taxes on labor income, the government can transfer resources from poor to rich in the first period, and by raising taxes on capital income, the government can transfer part of the superior returns

[18] This is formally shown in Appendix C.

back from the rich to the poor. Thus, by shifting the tax burden from labor to capital income, the government indirectly implements the missing transactions, thereby alleviating the market failure.

An alternative but ultimately equivalent interpretation of the positive capital tax is the following. Richer individuals obtain higher rates of return and therefore discount the future more heavily. As a result, redistribution from rich to poor is most efficient if it takes place relatively late in life. The easiest way to see this is to ignore behavioral responses to taxation and consider two individuals: a poor individual with a zero marginal rate of return ($y_a = 0$) and a rich individual with a positive rate of return ($y_a > 0$). The poor individual is indifferent between receiving one additional unit of resources in the first or the second period. However, the rich would rather lose a unit of resources in the second period than in the first period. Thus, to provide the same utility gains to the poor, the government imposes smaller utility losses on the rich by taxing them in the second period rather than the first period. In other words, redistribution of second-period resources Pareto dominates redistribution of first-period resources for the same distribution of income. With endogenous savings decisions, taxes on capital income optimally trade off the efficiency gains of redistributing late in life with the efficiency losses from distorting savings behavior.

To the best of our knowledge, this justification for positive taxes on capital income is entirely novel. For example, Gahvari and Micheletto (2016) explicitly state that taxes on capital income are redundant if the rich earn higher returns simply because they are rich and not because they have higher ability ($y_{aa} > 0$ and $y_n = 0$ in our terminology). In other words, they find no role for taxes on capital income if return heterogeneity stems from economies of scale. This follows directly from their assumption that both taxes on labor income and capital income are levied in the same period. Proposition 3 shows that their result breaks down if taxes on capital income are levied later in life than taxes on labor income. In that case, the government should tax capital income if rates of return are increasing in savings.[19]

One attractive feature of the optimality condition in eq. (27) is that it expresses the optimal tax on capital income exclusively in terms of sufficient statistics with clear empirical counterparts. The conditional elasticity of capital income could be estimated by using exogenous variation in marginal tax rates on capital income – even though currently empirical evidence on this particular elasticity is still scarce.[20] The distribution of capital income – and hence its hazard rate – can typically be obtained directly from administrative data. Moreover, there is an increasing amount of evidence on how rates of return vary with wealth (e.g., Fagereng et al., 2020). The only empirical matter on which we lack a good answer, is the extent to which heterogeneity in rates of return originates from scale economies in wealth management rather than other sources of heterogeneity such as type-dependent returns – as studied in Proposition 2.

We can use eq. (27) to determine the optimal tax rate on capital income at the top, provided the right tail of the distribution of capital income follows a Pareto distribution. This is shown in the following Corollary.

Corollary 1. *Assume that capital income is convex in wealth but not directly affected by ability ($y_{aa} \geq 0$*

[19]Naturally, our argument rests on the implicit assumption that taxes on labor income are levied when labor income is earned and cannot easily be deferred to later periods. Erosa and Gervais (2002) and Conesa, Kitao, and Krueger (2009) also propose positive taxes on capital income if taxes on labor income cannot be conditioned on age. However, their reasoning is very different from ours. In their model, future consumption is more complementary to leisure than current consumption. Hence, taxes on capital income are desirable to reduce distortions in labor supply (Corlett and Hague, 1953; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018). This argument does not play a role in our model, because preferences are weakly separable, so that there is no reason to tax capital income to alleviate distortions in labor supply (Atkinson and Stiglitz, 1976).

[20]This holds especially when compared to the abundance of estimates of the elasticity of labor income. See Seim (2017), Zoutman (2018), and Jakobsen et al. (2020) for some recent studies on the elasticity of wealth.

and $y_n = 0$ for all individuals). If the right tail of the distribution of capital income follows a Pareto distribution, and if the conditional elasticity of capital income and the marginal rate of return on savings converge to the constants $\hat{e}_{y|t}$ and $\hat{y}_a$ for high levels of income, then the optimal tax rate on capital income at the top of the income distribution is constant and given by:

$$\tau'(\hat{y}) = \frac{1}{\hat{e}_{y|z}} \frac{1}{p} \left(1 - \frac{\bar{y}_a}{\hat{y}_a}\right), \tag{28}$$

where a 'hat' denotes variables that refer to individuals in the top of the distribution of capital income, and $p = (1 - G(\hat{y}))/(\hat{y}g(\hat{y}))$ is the Pareto parameter of the right tail of the distribution of capital income.

Proof. Substituting $y_a = \bar{y}_a^+(y) = \hat{y}_a$, $e_{y|z} = \hat{e}_{y|z}$, and $yg(y)/(1 - G(y)) = p$ into eq. (27) yields eq. (28). □

The Pareto parameter p is an indication of the thinness of the tail of the capital-income distribution. Thus, the optimal top tax rate on capital income is decreasing in the elasticity of capital income at the top, $\hat{e}_{y|z}$, and the thinness of the income distribution's tail, p. Furthermore, it is increasing in the marginal rate of return on savings at the top relative to the average marginal rate of return.

Eq. (28) allows us to make a back-of-the-envelope calculation of optimal top tax rates on capital income. Based on the data in Fagereng et al. (2020), we assume that the average rate of return is about 1.1%, whereas the (risk-adjusted) rate of return of the wealthiest decile is about 0.5 percentage points higher. We compound these returns over a period of 32 years to accommodate for the 2-period life-cycle structure of our model – see the next Section for more details. This implies a value for $\bar{y}_a/\hat{y}_a$ of $0.42/0.66 = 0.63$.[21] Provided that the right tail of the distribution of capital income is comparable to that of wealth itself, a reasonable estimate for the Pareto parameter of the US distribution of capital income is $p = 1.6$ (e.g. Vermeulen, 2018). Unfortunately, there are few good estimates on the elasticity of capital income. Nevertheless, a reasonable estimate may be $\hat{e}_{y|z} = 1$, which is much higher than the elasticities of wealth found by Seim (2017) but smaller than those found by Zoutman (2018) and Jakobsen et al. (2020).[22] In the numerical simulation of our model (next Section), we target the more commonly estimated elasticity of intertemporal substitution, which also results in an elasticity of capital income of approximately 1. Taken together, these values imply an optimal top tax rate on capital income of 23 percent. This tax rate is substantial and can be much higher if smaller elasticities of capital income are assumed.

5.6.3 The general case

Propositions 2 and 3 present optimal taxes on capital income for two specific assumptions on the savings technology, which are consistent with two plausible microfoundations. We can also derive an optimality condition for the general case in which capital income is some function of savings and ability, $y^n = y(a^n, n)$, without imposing additional restrictions on the functional form of $y(\cdot)$. For this, we first write R^n, which is one plus the after-tax rate of return, as a function of capital income and ability only.

[21] The decision whether to compound returns has little bearing on our results. Without compounding, this ratio would equal $0.011/0.016 = 0.69$.

[22] After correcting for the base of the estimated elasticities (e.g., the net-of-tax rate or one plus the after-tax rate of return), the estimated elasticities range from values as small as 0.18 (Seim, 2017) to values that could well exceed 10 (Zoutman, 2018). Adding to the uncertainty around the empirical magnitude of this elasticity is the fact that these estimates concern elasticities of *wealth*, which may differ from elasticities of capital income in case of scale effects in returns.

To do so, we invert $y^n = y(a^n, n)$ to write $a^n = a(y^n, n)$ with $a_y = 1/y_a$ and $a_n = -y_n/y_a$. Substituting this back into R^n we obtain:

$$R^n = R(y^n, n) = 1 + (1 - \tau'(y^n))y_a(a(y^n, n), n). \tag{29}$$

Using this notation, we now present an expression for the optimal tax on capital income if capital income is a general function of savings and ability.

Proposition 4. *If capital income is a general function of savings and ability $y(a^n, n)$, the optimal marginal tax rate on capital income for every level of capital income y is implied by:*

$$\begin{aligned}\left(\frac{\tau'(y)y_a}{\lambda_1/\lambda_2}\right)e_{y|z} = &\left(\frac{T'(z)}{1-T'(z)} + \frac{sy_a\tau'(y)}{\lambda_1/\lambda_2}\right)e_z\left(\frac{\xi_{y|z}}{\xi_z}\right)\\ &+ \frac{1-G(y)}{yg(y)}\left(\frac{1+\bar{y}_a^+(y) - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right)\\ &- \frac{1}{yg(y)}\int_y^\infty\left(\frac{\partial R(y^n,n)}{\partial n}\right)\left(\frac{\mathrm{d}y^n}{\mathrm{d}n}\right)^{-1}E_\tau^n g(y^n)\mathrm{d}y^n.\end{aligned} \tag{30}$$

Proof. See Appendix C. □

The left-hand side and the first two terms on the right-hand side of eq. (30) are familiar from Propositions 2 and 3. The left-hand side represents the savings distortions associated with an increase in the tax on capital income. The first term on the right-hand side is the degree to which taxes on capital income distort labor supply less than taxes on labor income – for the same amount of income redistribution. And the second term on the right-hand side reflects the role of taxes on capital income in correcting market failures associated with scale effects. Thus, only the third term on the right-hand side of eq. (30) is new.

Unfortunately, the intuition behind the last term in eq. (30) is not straightforward. The term originates from the fact that we compare an increase in the marginal tax on labor income around z with a distributionally equivalent increase in marginal taxes on capital income. Hence, both reforms cause the same utility losses for individuals with income above z. If the discount rate is increasing with ability ($\partial R(y^n, n)/\partial n > 0$), then the increase in the capital tax liability should be increasing with ability to obtain the distributional equivalent of the increase in the marginal tax on labor income. This can be achieved only if marginal taxes on capital income are raised for *all* levels of capital income above y for which $\partial R(y^n, n)/\partial n > 0$. These additional increases in marginal taxes yield additional distortions that are proportional to the marginal excess burden of the tax on capital income E_τ. The final term in eq. (30) is negative if $\partial R(y^n, n)/\partial n > 0$ and reflects this additional distortion. Naturally, the reverse holds if $\partial R(y^n, n)/\partial n < 0$, in which case the final term is positive. Hence, its sign is *a priori* ambiguous and depends on whether the marginal rate of return is increasing or decreasing in ability for a given capital income ($\partial R(y^n, n)/\partial n \gtrless 0$).

We *are* able to sign the term if return heterogeneity stems from a convex combination of type and scale dependence, i.e., if $y_n > 0$ and $y_{na} = 0$ as in our first microfoundation, and $y_{aa} > 0$ as in our second microfoundation. To see this, take the derivative of eq. (29) to find:

$$\frac{\partial R(y^n, n)}{\partial n} = (1 - \tau')\left(y_{an} - \frac{y_{aa}y_n}{y_a}\right). \tag{31}$$

Substituting for $y_n > 0$, $y_{na} = 0$, and $y_{aa} > 0$ yields $\partial R(y^n, n)/\partial n < 0$. Intuitively, conditional on capital income, higher ability generates more income from the closely-held asset so that an individual needs to invest less in the common savings technology with scale benefits to obtain a given amount of capital income. As a result, for given capital income, higher ability is associated with a *lower* marginal rate of return $R(y^n, n)$. Substitute this back into eq. (30) to find that the final term on the right-hand side is positive. Hence, the optimal tax on capital income is unambiguously positive if return heterogeneity is caused by a combination of the two microfoundations that we have discussed above.

6 Numerical simulation

In this Section, we simulate optimal non-linear taxes on labor and capital income in the US. We focus on the case where heterogeneity in returns on capital originates from closely-held assets.

6.1 Calibration and simulation method

Our simulations determine optimal marginal tax rates on labor and capital income exclusively in terms of structural model primitives: the utility function, the social welfare function, the ability distribution, and the distribution of capital returns. The reason for doing so is twofold. First, and most importantly, while our optimal tax schedules above are expressed in terms of sufficient statistics, we currently lack the estimates of the elasticity of capital income with respect to wages per hour worked conditional on labor income. Our structural model allows us to map measured return heterogeneity into this conditional elasticity. Second, the elasticities and the distributions of labor and capital income are endogenous with respect to the tax schedules. Biases in optimal tax calculations can appear if optimal taxes are calculated using only sufficient statistics, and optimal schedules deviate substantially from current schedules (Chetty, 2009).

We calibrate the ability distribution in our model such that the resulting income distribution matches the empirical income distribution in the US. We calibrate heterogeneous rates of return on the basis of Norwegian estimates from Fagereng et al. (2020). We use Norwegian estimates because we are not aware of good measures of return heterogeneity for the US. The values reported by Fagereng et al. (2020) are the most comprehensive measures currently available. This reliance on a combination of US and Norwegian data is in line with other recent studies that numerically simulate heterogeneous returns on capital (e.g., Guvenen et al., 2019). We calibrate the parameters of the utility function to match empirical estimates for the labor supply elasticity and the intertemporal elasticity of substitution in consumption. The following subsections explain the calibration of the ability distribution, the returns on capital, the utility function, and social preferences for income redistribution.

6.1.1 Timing and capital returns

One of the main challenges in our calibration is to map our two-period model to a real world in which income is earned and taxes are paid on an annual basis. We make the simplifying assumption that each of the two periods consist of 32 identical years. Individuals make one labor-supply decision and stick to this decision for the first 32 years of life – yielding the same z^n in each year. Moreover, individuals make one decision on how much of their after-tax labor income to save for retirement and also stick to these savings a^n for the first 32 years of life. Capital income is only taxed on realization, i.e., when assets are sold in the final 32 years of life. During that period, individuals are retired and simply consume their savings and their after-tax capital income. Because the two periods are of equal length and all years within a period are identical, it is *as if* an individual consumes in year k of the second period the savings and their after-tax returns from year k of the first period. This allows us to retain the two-period structure of our theoretical model while calibrating it with empirically observable annual data.

We focus our simulations on the first microfoundation with type-dependent returns on capital. Thus, we can parameterize the capital income function as follows:

$$y^n = ra^n + \theta(n), \tag{32}$$

where we refer to $\theta(n)$ as the *excess returns on capital.* Note that this is a restatement of eq. (12) with $\theta(n) \equiv \pi(b(n), n) - rb(n)$. All individuals obtain the same marginal rate of return r but differ in their average rates of return $r + \theta(n)/a^n$, such that $y_a = r$, $y_{aa} = 0$, $y_{an} = 0$, and $y_n = \theta' > 0$. As y^n refers to capital income in the second period and a^n to savings in the first period, we calibrate the 'lifetime' marginal rate of return r on the basis of a 'yearly' marginal rate of return r^{yearly}, compounded for 32 years:

$$r = (1 + r^{\text{yearly}})^{32} - 1. \tag{33}$$

We choose a value of the yearly marginal rate of return of $r^{\text{yearly}} = 0.03$. This is close to the average yearly return on a 30-year Treasury bill in the last 10 years, which was 3.28% (Federal Reserve Bank of St. Louis, 2020). By compounding this value for 32 years, we obtain a lifetime marginal rate of return of $r = 1.58$.

We denote the yearly *average* rate of return by $\bar{r}^{\text{yearly}}$. Unlike the marginal rate of return, $\bar{r}^{\text{yearly}}$ also accounts for the yearly inframarginal excess returns. We have evidence on the distribution of these yearly average rates of return from Fagereng et al. (2020). Thus, to calibrate the distribution of lifetime excess returns $\theta(n)$, we need to relate these lifetime excess returns to the yearly average rate of return. By definition, lifetime capital income y^n is obtained by compounding the average yearly return on savings for 32 years. We can therefore write:

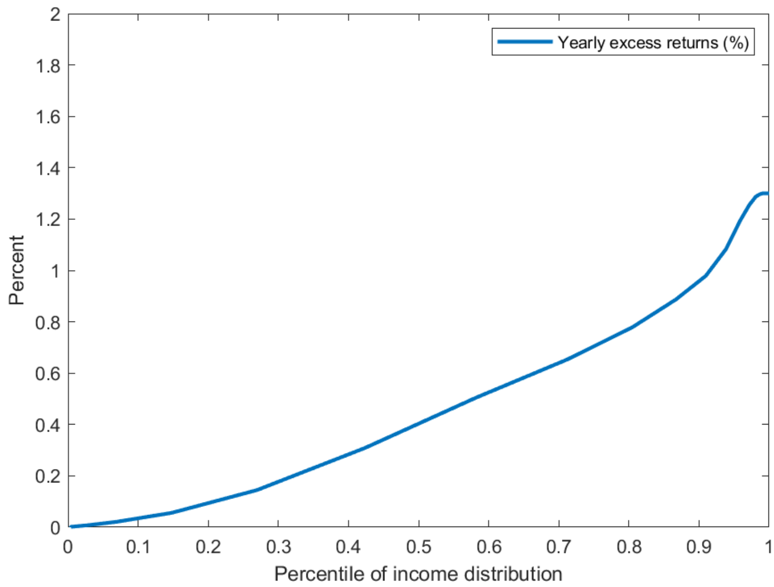
$$ra^n + \theta(n) = ((1 + \bar{r}^{\text{yearly}})^{32} - 1)a^n. \tag{34}$$

In our baseline calibration, we set excess returns of the poorest households to zero ($\theta(0) = 0$), implying that they only have access to the commonly traded asset with rate of return r. We calibrate the rest of the distribution of excess returns $\theta(n)$ in order to match the gradient of risk-adjusted yearly *average* rates of return over the wealth distribution as shown by Fagereng et al. (2020) in their Figure 3. They find that risk-adjusted average rates of return rise (at an increasing rate) by about 1.3 percentage points from the smallest to the largest wealth percentile. Thus, we calibrate $\theta(n)$ such that $\bar{r}^{\text{yearly}}$, as implied by eq. (34), increases by about 1.3 percentage points over the wealth distribution.[23] Figure 1 shows the calibrated yearly excess rates of return ($\bar{r}^{\text{yearly}} - r^{\text{yearly}}$) over the wealth distribution.

It is unclear to what extent we can generalize the findings of Fagereng et al. (2020) to other countries. For example, Bach, Calvet, and Sodini (2020) also document significant return heterogeneity in Sweden, but they argue that this can mostly be explained by differences in risk taking. Because, as of yet, there is no consensus on the quantitative importance of type-dependent returns, we also explore the sensitivity of our results to a conservative case, in which the heterogeneity in rates of return is only 20% of what is found by Fagereng et al. (2020). In that case, excess returns only increase by about 0.26 percentage

[23] In particular, we match estimates by Fagereng et al. (2020) for the 10th, 50th, 90th, and 99th percentiles of the wealth distribution and use a cubic interpolation for the remainder of the wealth distribution.

Figure 1: Yearly excess rate of return in the calibration



points over the income distribution.

6.1.2 Ability distribution

We calibrate the ability distribution to ensure that the simulated distribution of labor income roughly corresponds to the empirical income distribution in the US. We assume that ability n follows a log-normal distribution up to a certain level of ability n^*, after which it follows a Pareto distribution. Thus, $\ln n \sim N(\mu, \mathrm{sd})$ for $n \in (0, n^*)$, where mean μ and standard deviation sd of log ability are chosen such that the mean and the median of labor income match their 2018 US values of \$54.906 and \$40.453 (U.S. Census Bureau, 2019). We introduce a mass point of 0.4 percent of the population at the bottom of the ability distribution ($n = 0$). This is to ensure that optimal marginal tax rates are always positive at the bottom. Finally, we append the ability distribution with a Pareto tail for ability levels above n^*. The Pareto parameter of the right tail of the ability distribution is set to $p = 2.5$.[24] We choose n^* and the scale parameter of the Pareto tail such that the probability density function and its first derivative are continuous. This avoids irregularities in the optimal tax schedules.

6.1.3 Utility function and behavioral elasticities

We adopt a utility function that is iso-elastic in both consumption and labor supply:

$$U^n = \frac{c_1^{1-1/\sigma}}{1-1/\sigma} + \beta\frac{c_2^{1-1/\sigma}}{1-1/\sigma} - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}, \quad \beta, \varepsilon, \sigma > 0, \tag{35}$$

where σ is the elasticity of intertemporal substitution in consumption, ε is the Frisch elasticity of labor supply, and β captures the time preference of individuals. We set the rate of time preference at $\beta = 1/(1+r)$. Similar to Saez (2001) and in line with empirical estimates discussed in Blundell and MaCurdy (1999) and Meghir and Phillips (2010), we adopt a Frisch elasticity of labor supply ε equal to 0.20.

[24] We use a somewhat higher value for the Pareto parameter of labor income than conventional estimates of around 2, since the latter apply to total income (Atkinson, Piketty, and Saez, 2011). Since labor income is more equally distributed than capital income, the Pareto parameter for labor income is likely a bit higher.

In our baseline scenario, the elasticity of intertemporal substitution (EIS) is set to $\sigma = 1.5$. This is considerably higher than most of the estimates in the literature on intertemporal choice.[25] There are two main reasons why we choose a relatively high value for the EIS. First, our analysis does not incorporate any behavioral responses of the capital tax base other than intertemporal substitution of consumption, such as entrepreneurial effort, international capital mobility, and tax evasion. As a result, an empirically realistic value of the EIS may still yield unrealistically low values for the elasticity of taxable capital income. We therefore prefer to err on the conservative side by choosing a relatively high value for the EIS.[26] Second, a low value of the EIS, combined with positive returns to ability, may generate such strong income effects that both labor supply and labor earnings are backward bending in ability. Consequently, the labor tax base would no longer be monotonic in n, which would violate the second-order conditions of the optimal tax problem. This latter reason is much less relevant if excess returns are relatively small. Therefore, we do allow for a lower EIS of $\sigma = 0.5$ when we calibrate return heterogeneity at 20% of our baseline calibration.

6.1.4 Social welfare function

The social welfare function is assumed to be exponential:

$$\mathcal{W} = -\int_0^\infty \frac{\exp(-U^n\gamma)}{\gamma} f(n)\mathrm{d}n, \tag{36}$$

where $\gamma > 0$ is a parameter that determines the degree of inequality aversion. If $\gamma \to 0$, the government is utilitarian. If $\gamma \to \infty$, the government is Rawlsian. In the baseline, we set γ to ensure that the optimal tax on labor income is comparable to Saez (2001). In particular, the average income-weighted marginal labor income tax is 57% and the asymptotic top rate is 74% compared to 59% and 81% in Saez (2001, Table 2). We also provide sensitivity analyses for alternative values of γ.

6.1.5 Calibration

Similar to Saez (2001), we employ a linear approximation of the US tax system when calibrating the distributions of ability and capital returns. The marginal tax rate on labor income is set to 0.35, close to the approximation of 0.34 found by Heathcote, Storesletten, and Violante (2017). The marginal tax rate on capital income is set to 0.15, the middle bracket tax rate of the tax on long-term capital gains in the United States. The intercept of the labor income tax is \$10,000 and the intercept of the capital tax schedule is normalized to zero.

To calibrate the distribution of ability and capital returns we proceed as follows. We use a nested fixed-point algorithm to find the parameters of the skill distribution, μ and sd, and the distribution of the excess returns, $\theta(n)$. We start with an initial guess for μ and sd, and set excess returns to zero for all individuals. Based on the elasticities we imposed on the utility function, we are then able to simulate individual behavior, which yields values of both savings a^n and labor supply l^n. Given this behavior, we update excess returns $\theta(n)$ in the inner loop of our algorithm to match the return heterogeneity found by Fagereng et al. (2020). We simulate individual behavior again with the updated excess returns, reiterating this process until it converges to a fixed point for excess returns. Given labor supply l^n, we

[25] Empirical estimates range from 0 to 2 but the vast majority of the estimates are well below 1. For reviews of the literature, see Attanasio and Weber (2010) and Best et al. (2020). For a meta study, see Havránek (2015).

[26] Unfortunately, we do not know much empirically about the elasticity of capital income with respect to the net interest rate, $e^n_{y|z}$. In our baseline calibration this elasticity is approximately 1, which is within a very wide range of empirical estimates. See also footnote 22

then update μ and sd in the outer loop of our algorithm so that mean and median labor income matches their US counterparts. We then reiterate the whole process until it converges to a fixed point for both the distribution of excess returns and the parameters of the ability distribution. Finally, we choose the exogenous revenue requirements in both periods such that the government budget constraints in eqs. (16) and (17) bind.

6.1.6 Simulation method

Optimal tax rates on labor income and capital income are given by eqs. (25) and (26) in Propositions 1 and 2. For simulation purposes, these optimal tax conditions are completely rewritten in terms of model primitives. This involves a fair amount of algebra and is relegated to Appendix D. We obtain a numerical approximation of the optimal non-linear tax schedules by using a fixed-point algorithm similar to the one developed by Mankiw, Weinzierl, and Yagan (2011). In particular, we start with an initial guess of the optimal tax schedules. We then simulate individual behavior with these tax schedules. Next, we impose the first-order conditions for the optimal tax schedules to update the initial guess of the tax schedules and adjust the intercepts by imposing the government's budget constraints of eqs. (16) and (17). We reiterate this process until the optimal tax schedules converge to a fixed point.

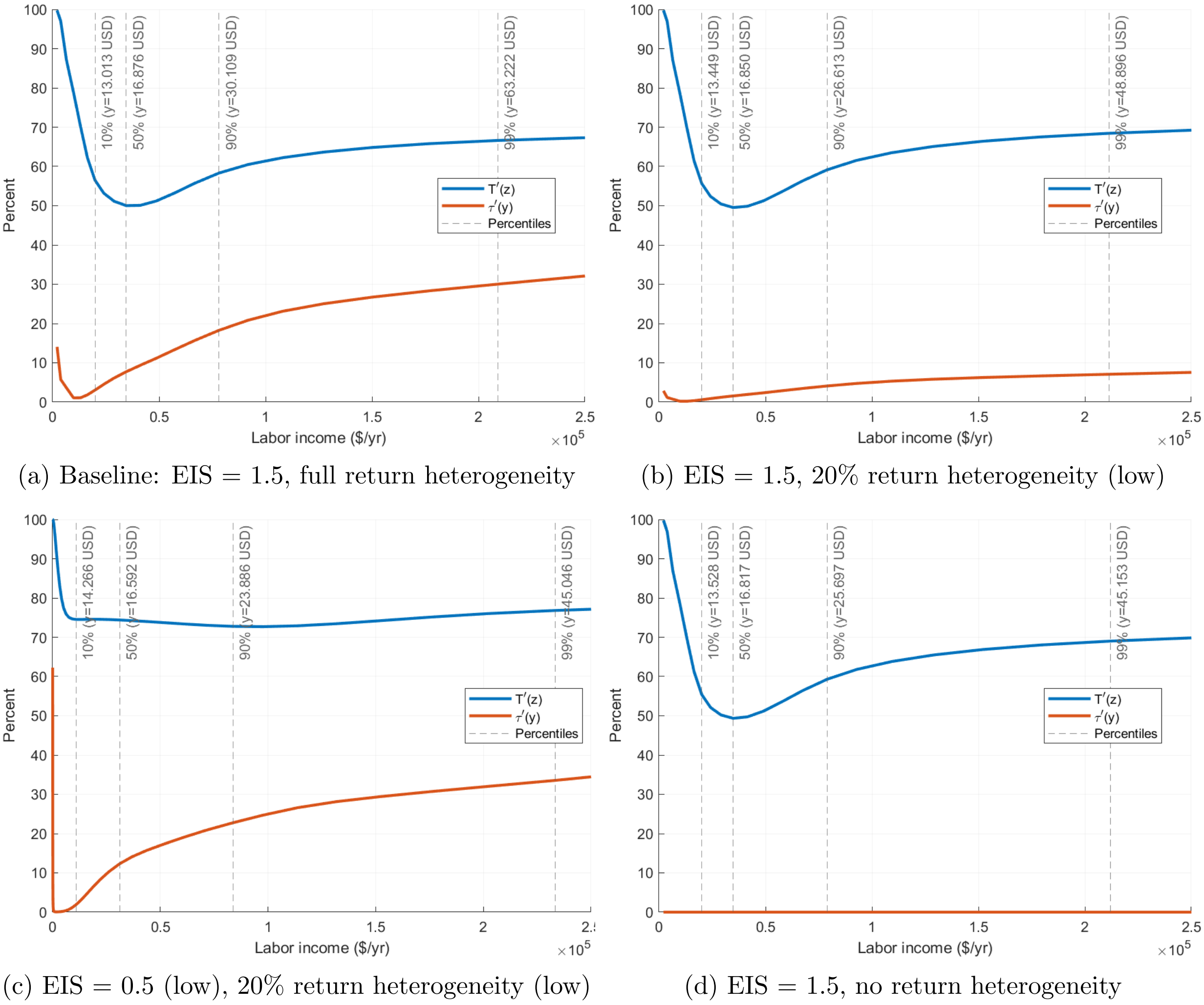
6.2 Optimal income taxes

We present optimal tax rates on labor income and capital income in Figure 2. We do so for four different scenarios, reflecting the substantial empirical uncertainty about the extent of return heterogeneity and the elasticity of capital income. In each of the four panels, the blue line reflects the optimal marginal tax rates on labor income and the red line reflects the optimal marginal tax rates on capital income. Both tax schedules are plotted against labor income on the horizontal axis, so that marginal tax rates on both labor and capital income can be inferred for each individual, and income percentiles can be indicated with the same dashed vertical lines. For the 10th, 50th, 90th, and 99th percentiles, we give the associated level of capital income within parentheses.[27]

Panel (a) covers our baseline simulation, where the elasticity of intertemporal substitution (EIS) is set at $\sigma = 1.5$, and return heterogeneity is calibrated to the full magnitude found in Fagereng et al. (2020). Panels (b) and (c) depict optimal tax schedules under alternative assumptions on return heterogeneity and the elasticity of intertemporal elasticity. Panel (b) considers the case in which the extent of return heterogeneity is only 20% of what we assume in the baseline. Panel (c) retains the reduced extent of return heterogeneity but now reduces the EIS to a more conventional level of $\sigma = 0.5$. Together, the first three panels provide a reasonable range for optimal marginal tax rates on capital income. Moreover, they allow us to explore the sensitivity of our numeric results to variations in return heterogeneity and the EIS. Finally, Panel (d) covers the case without return heterogeneity. The analysis in Panel (d) otherwise employs the same parameters as in the baseline and confirms the theory by featuring zero optimal tax rates on capital income. It also corresponds most closely to the optimal non-linear income tax of Saez (2001), where all heterogeneity is in labor income.

[27] These levels of capital income cannot readily be compared to conventional statistics on the distribution of capital income. This is because we define capital income as the compounded returns on one year's worth of savings while available statistics normally define it as the yearly returns on accumulated savings from all previous years. Moreover, capital income in our model depends on the arbitrary assumption that only the intercept of our labor income tax can be negative. This latter assumption is irrelevant for our results but does affect household savings and thus levels of capital income.

Figure 2: Optimal marginal capital and labor income taxes



(a) Baseline: EIS = 1.5, full return heterogeneity

(b) EIS = 1.5, 20% return heterogeneity (low)

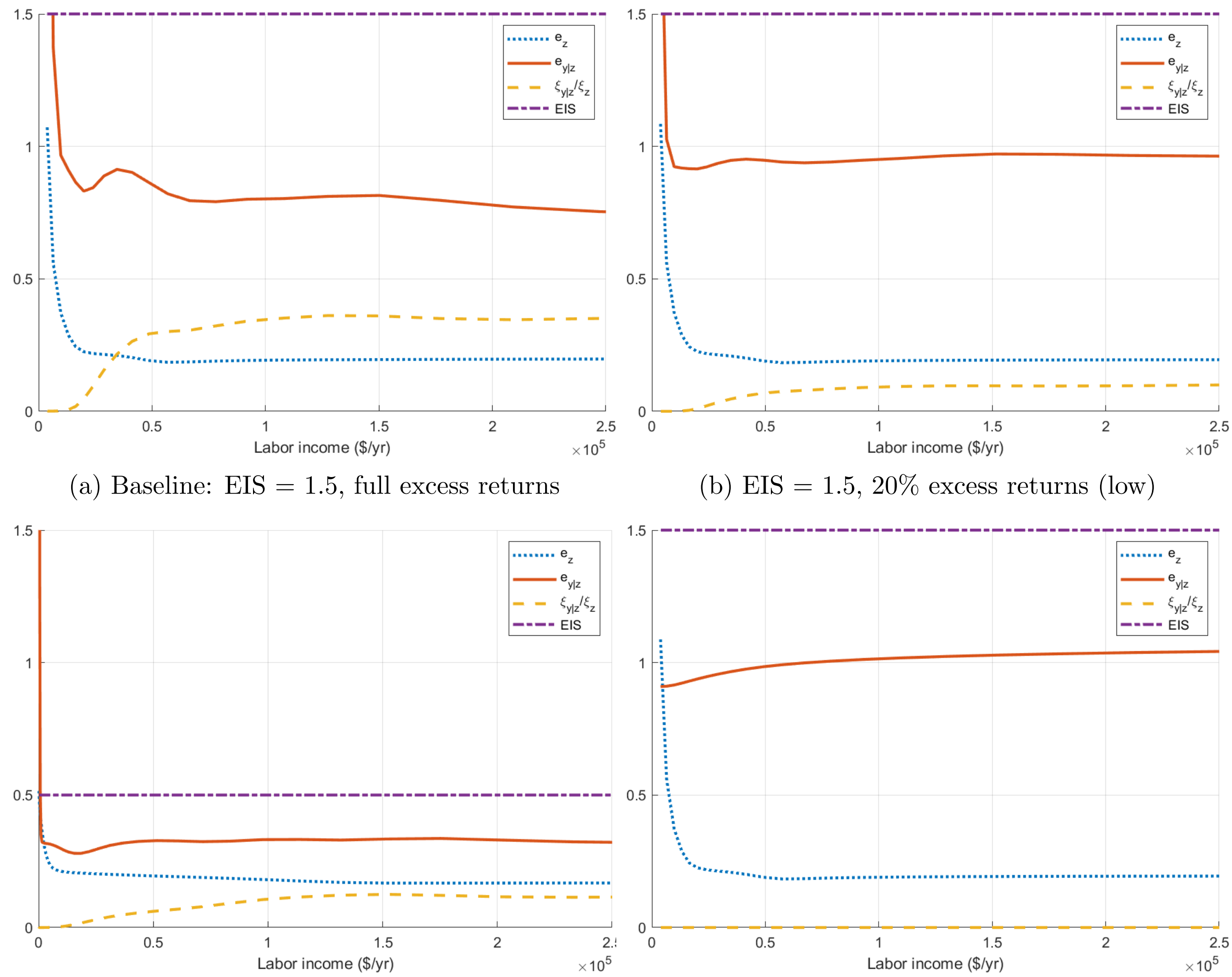
(c) EIS = 0.5 (low), 20% return heterogeneity (low)

(d) EIS = 1.5, no return heterogeneity

Figure 2 illustrates two main insights from our numerical simulations. The first main insight is that optimal marginal tax rates on capital income are positive and economically significant for a majority of taxpayers in all but the most conservative of our simulations. The income-weighted average of the optimal marginal tax on capital income is equal to 14 percent in the baseline simulation of Panel (a) of Figure 2. It is 16 percent in the case with a more reasonable EIS of 0.5 but a very small amount of return heterogeneity as in Panel (c). These are economically significant tax rates, particularly considering the fact that our model disregards all other relevant reasons to tax capital as surveyed in Section 2. Only in the most conservative of our simulations – if we assume both a very small amount of return heterogeneity *and* strong intertemporal consumption responses – as in Panel (b), we find a small income-weighted average optimal marginal tax on capital income of about 3 percent.

The quantitative importance of the optimal tax on capital income is even more apparent at the top of the income distribution. The average income-weighted marginal tax rate for the top decile is equal to 26 percent in Panel (a) of Figure 2 and ranges from 6 percent in Panel (b) to 30 percent in Panel (c). Similarly, the optimal tax rate at the 99th percentile of the income distribution ranges from 7 to 33 percent.

Figure 3: Elasticities under optimal taxation.



(a) Baseline: EIS = 1.5, full excess returns

(b) EIS = 1.5, 20% excess returns (low)

(c) EIS = 0.5 (low), 20% excess returns (low)

(d) Atkinson-Stiglitz: EIS = 1.5, no excess returns

The second main insight from our simulations is that the optimal marginal tax rates on capital income are increasing in income for nearly the entire income distribution. While, strictly speaking, the optimal tax schedule is U-shaped in income, the downward-sloping part of the schedule is very local and only affects the few smallest percentiles. Thus, broadly speaking, our simulations suggest that marginal tax rates on capital income should rise with capital income.

The shape of the optimal tax schedule is mainly driven by two factors represented in the optimal tax formula of eq. (26). First, the optimal marginal tax rate on capital income is increasing in the ratio of ability elasticities $\xi_{y|z}/\xi_z$, which captures the informational content of the capital tax base relative to the labor tax base. Indeed, Figure 3 shows that our calibration of return heterogeneity implies that this ratio is increasing over the entire income distribution. Second, the optimal marginal tax rate on capital income is increasing in the tax wedge on labor income $T'(z)/(1-T'(z))$. This explains the U-shape in the capital tax schedule as the optimal tax rate on labor income $T'(z)$ comes close to 100 percent as income goes to zero. However, in all of our simulations, the upward sloping $\xi_{y|z}/\xi_z$ offsets most of the downward sloping part of the tax wedge on labor income.[28]

While both insights are robust across simulations, the precise magnitude of the optimal marginal tax rates on capital income is sensitive to assumptions on return heterogeneity and the intertemporal elasticity of substitution in consumption. The comparison of Panels (a) and (b) of Figure 2 shows that optimal marginal tax rates on capital income decline if return heterogeneity is only a fifth of that in the baseline. In line with this finding, Figure 3 shows that the ratio of ability elasticities $\xi_{y|z}/\xi_z$ is significantly reduced in Panel (b). Similarly, the comparison of Panels (b) and (c) of Figure 2 demonstrates that optimal tax rates rise significantly if the EIS declines from 1.5 to 0.5, since this reduces the conditional elasticity of capital income $e_{y|z}$, as shown in Figure 3. The robustness analyses in Appendix E show that the optimal tax rates on both labor and capital income are lower if the government is utilitarian ($\gamma = 0$) or if the Frisch elasticity of labor supply is higher ($\varepsilon = 0.5$).

The sensitivity of our results, together with the stylized nature of our model, gives reason to interpret the numerical results with caution. The simulations cannot provide a narrow range for optimal marginal taxes on capital income. However, the results do suggest that optimal marginal tax rates on capital income are positive and increasing with income, and that they are likely to be quantitatively significant for a large part of the income distribution.

7 Conclusion

A large body of empirical evidence demonstrates that people differ in the rates of return on their capital, even after controlling for risk-taking behavior. This paper analyzes optimal non-linear taxes on labor and capital income if individuals have heterogeneous capital returns. We show that optimal taxes on capital income are positive both if return differences originate from closely-held assets and if they originate from scale effects in wealth management. An empirically plausible numerical calibration of our model indicates that the optimal non-linear tax on capital income may be significant in magnitude and increasing in income.

Future research may extend the current paper in a number of directions. First, it would be interesting to add idiosyncratic and systematic risk and portfolio choice to analyze optimal taxes on capital income with heterogeneous returns. If portfolio choice is correlated with earning ability, it might be optimal to distort risk-taking behavior for income redistribution. Second, this paper's two-period model structure might be extended to allow for a multiple-period life-cycle or OLG model to explore the

[28] Figure 3 also shows that the elasticities e_z and $e_{y|z}$ spike around the bottom of the income distribution. This is because of a highly negative second-order derivative of the labor tax function, $T''(z)$. However, these spikes cannot explain the shape of the optimal capital tax schedule because the ratio of the two elasticities, $e_z/e_{y|z}$ – which is what matters according to eq. (26) – is relatively stable.

quantitative robustness of optimal taxes on capital income in more realistic multiple-period models. Third, the model could be extended with entrepreneurial effort, tax avoidance in capital taxes and cross-border shopping. Such settings would allow for an expansion of the government instrument set to include wealth and consumption taxes.

Our paper contributes to a large and growing literature that casts doubt on the relevance of the zero capital tax results (Atkinson and Stiglitz, 1976; Chamley, 1986; Judd, 1985). Our paper provides an important argument why the zero tax results offer little practical guidance for applied tax policy. Capital income should be taxed if individuals differ in the rates of return to capital, for which there is overwhelming and well-documented evidence. Hence, even if we ignore all other relevant reasons to tax capital income, the policy implications of our paper are clear: optimal taxes on capital income are positive, economically significant, and increasing with capital income.

References

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson. 2009. "The Age of Reason: Financial Decisions Over the Life Cycle and Implications for Regulation." *Brookings Papers on Economic Activity* 2009 (2):51–117.

Aiyagari, S. Rao. 1995. "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints and Constant Discounting." *Journal of Political Economy* 103 (6):1158–1175.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. *World Inequality Report 2018*. World Inequality Lab.

Arfken, George B. and Hans J. Weber. 2005. *Mathematical Methods for Physicists International Student Edition*. Elsevier, Burlington, MA.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49 (1):3–71.

Atkinson, Anthony B. and Agnar Sandmo. 1980. "Welfare Implications of the Taxation of Savings." *Economic Journal* 90 (359):529–549.

Atkinson, Anthony B. and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct Versus Indirect Taxation." *Journal of Public Economics* 6 (1-2):55–75.

Attanasio, Orazio P. and Guilermo Weber. 2010. "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy." *Journal of Economic Literature* 48 (3):693–751.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2020. "Rich Pickings? Risk, Return, and Skill in Household Wealth." *American Economic Review* forthcoming.

Barber, Brad M., Terrance Odean, and Lu Zheng. 2005. "Out of Sight, Out of Mind: the Effects of Expenses on Mutual Fund Flows." *Journal of Business* 78 (6):2095–2120.

Bastani, Spencer and Daniel Waldenström. 2020. "How Should Capital be Taxed?" *Journal of Economic Surveys* forthcoming.

Benartzi, Shlomo and Richard H. Thaler. 2001. "Naive Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review* 91 (1):79–98.

Benhabib, Jess, Alberto Bisin, and Mi Luo. 2019. "Wealth Distribution and Social Mobility in the US: A Quantitative Approach." *American Economic Review* 109 (5):1623–47.

Best, Michael Carlos, James S. Cloyne, Ethan Ilzetzki, and Henrik J. Kleven. 2020. "Estimating the elasticity of intertemporal substitution using mortgage notches." *Review of Economic Studies* 87 (2):656–690.

Blundell, Richard and Thomas MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*, vol. 3, edited by Orley C. Ashenfelter and David Card. Amsterdam, the Netherlands: Elsevier, 1559–1695.

Boháček, Radim and Michal Kejak. 2018. "Optimal Government Policies in Models with Heterogeneous Agents." *Journal of Economic Theory* 176:834–858.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115 (5):707–747.

Campbell, John Y. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review: Papers & Proceedings* 106 (5):1–30.

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

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-form Methods." *Annual Review of Economics* 1:451–488.

Choi, James J., David Laibson, and Brigitte C. Madrian. 2005. "Are Empowerment and Education Enough? Underdiversification in 401 (k) Plans." *Brookings Papers on Economic Activity* 2005 (2):151–213.

———. 2010. "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds." *Review of Financial Studies* 23 (4):1405–1432.

———. 2011. "$100 Bills On the Sidewalk: Suboptimal Investment in 401 (k) Plans." *Review of Economics and Statistics* 93 (3):748–763.

Christiansen, Vidar. 1981. "Evaluation of Public Projects under Optimal Taxation." *Review of Economic Studies* 48 (3):447–457.

———. 1984. "Which Commodity Taxes Should Supplement the Income Tax?" *Journal of Public Economics* 24 (2):195–220.

———. 1993. "A Normative Analysis of Capital Income Taxes in the Presence of Aggregate Risk." *Geneva Papers on Risk and Insurance Theory* 18 (1):55–76.

Christiansen, Vidar and Matti Tuomala. 2008. "On Taxing Capital Income with Income Shifting." *International Tax and Public Finance* 15 (4):527–545.

Conesa, Juan C., Sagiri Kitao, and Dirk Krueger. 2009. "Taxing Capital? Not a Bad Idea After All!" *American Economic Review* 99 (1):25–48.

Corlett, Wilfred J. and Douglas C. Hague. 1953. "Complementarity and the Excess Burden of Taxation." *Review of Economic Studies* 21 (1):21–30.

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet. 2001. "Direct versus Indirect Taxation: The Design of Tax Structure Revisited." *International Economic Review* 42 (3):781–799.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1):83–95.

Diamond, Peter A. and James A. Mirrlees. 1978. "A Model of Social Insurance with Variable Retirement." *Journal of Public Economics* 10 (3):295–336.

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

Diamond, Peter A. and Johannes Spinnewijn. 2011. "Capital Income Taxes with Heterogeneous Discount Rates." *American Economic Journal: Economic Policy* 3 (4):52–76.

Egan, Mark, Gregor Matvos, and Amit Seru. 2018. "The Market for Financial Adviser Misconduct." *Journal of Political Economy* 127 (1):233–295.

Erosa, Andrès and Martin Gervais. 2002. "Optimal Taxation in Life-Cycle Economies." *Journal of Economic Theory* 105 (2):338–369.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. "Heterogeneity and persistence in returns to wealth." *Econometrica* 88 (1):115–170.

Federal Reserve Bank of St. Louis. 2020. "30-Year Treasury Constant Maturity Rate, Percent, Annual, Not Seasonally Adjusted." Federal Reserve Economic Data, Federal Reserve Bank of St. Louis.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6):2071–2111.

Gahvari, Firouz and Luca Micheletto. 2016. "Capital Income Taxation and the Atkinson–Stiglitz Theorem." *Economics Letters* 147:86–89.

Gerritsen, Aart. 2016. "Optimal Taxation when People Do Not Maximize Well-being." *Journal of Public Economics* 144:122–139.

Goetzmann, William N. and Alok Kumar. 2008. "Equity Portfolio Diversification." *Review of Finance* 12 (3):433–463.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies* 70 (3):569–587.

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin. 2014. "A Variational Approach to the Analysis of Tax Systems." Working Paper 20780, National Bureau of Economic Research, Cambridge, MA.

Gordon, Roger H. 1985. "Taxation of Corporate Capital Income: Tax Revenues Versus Tax Distortions." *Quarterly Journal of Economics* 100 (1):1–27.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen. 2019. "Use It or Lose It: Efficiency Gains from Wealth Taxation." Working Paper 26284, National Bureau of Economic Research, Cambridge, MA.

Havránek, Tomáš. 2015. "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting." *Journal of the European Economic Association* 13:1180–1204.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2017. "Optimal Tax Progressivity: An Analytical Framework." *Quarterly Journal of Economics* 132 (4):1693–1754.

Hubbard, R. Glenn and Kenneth L. Judd. 1986. "Liquidity Constraints, Fiscal Policy, and Consumption." *Brookings Papers on Economic Activity* 1986 (1):1–59.

Immervoll, Herwig, Henrik J. Kleven, Claus T. Kreiner, and Emmanuel Saez. 2007. "Welfare Reform in European Countries: A Microsimulation Analysis." *Economic Journal* 117 (516):1–44.

Jacobs, Bas. 2013. "From Optimal Tax Theory to Applied Tax Policy." *FinanzArchiv* 69 (3):338–389.

Jacobs, Bas and Robin Boadway. 2014. "Optimal Linear Commodity Taxation under Optimal Non-Linear Income Taxation." *Journal of Public Economics* 117 (1):201–210.

Jacobs, Bas and Lans A. Bovenberg. 2010. "Human Capital and Optimal Positive Taxation of Capital Income." *International Tax and Public Finance* 17 (5):451–478.

Jacobs, Bas and Alexandra V. Rusu. 2018. "Why is the Long-Run Tax on Capital Income Zero? Explaining the Chamley-Judd Result." Mimeo, Erasmus University Rotterdam.

Jacobs, Bas and Dirk Schindler. 2012. "On the Desirability of Taxing Capital Income in Optimal Social Insurance." *Journal of Public Economics* 96 (9–10):853–868.

Jacquet, Laurence and Etienne Lehmann. 2017. "Optimal Income Taxation with Composition Effects." Working Paper 6654, CESifo, Munich, Germany.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden. 2013. "Optimal Redistributive Taxation with Both Intensive and Extensive Responses." *Journal of Economic Theory* 148 (5):1770–1805.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. 2020. "Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark." *Quarterly Journal of Economics* 135 (1):329–388.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics* 28 (1):59–83.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens. 2018. "Investor Sophistication and Capital Income Inequality." *Journal of Monetary Economics* 107:18–31.

King, Mervyn. A. 1980. "Savings and Taxation." In *Public Policy and the Tax System*, edited by Gordon A. Hughes and Geoffrey M. Heal. London: Allen and Unwin, 1–35.

Kristjánsson, Arnaldur S. 2016. "Optimal Taxation with Endogenous Return to Capital." Mimeo, Oslo University.

Lehmann, Etienne, Sander Renes, Kevin Spiritus, and Floris T. Zoutman. 2018. "Optimal Joint Taxation." Mimeo, Erasmus University Rotterdam.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S. Mitchell. 2017. “Optimal Financial Knowledge and Wealth Inequality.” *Journal of Political Economy* 125 (2):431–477.

Lusardi, Annamaria and Olivia S. Mitchell. 2011. “Financial Literacy Around the World: An Overview.” *Journal of Pension Economics and Finance* 10 (4):497–508.

Mankiw, N. Gregory, Matthew C. Weinzierl, and Danny Yagan. 2011. “Optimal Taxation in Theory and Practice.” *Journal of Economic Perspectives* 23 (4):147–174.

Meghir, Costas and David Phillips. 2010. “Labour Supply and Taxes.” In *The Mirrlees Review. Dimensions of Tax Design*, edited by James A. Mirrlees et al. Oxford, United Kingdom: Oxford University Press, 202–274.

Mill, John S. 1848. *Principles of Political Economy.* London, United Kingdom: John W. Parker.

Mirrlees, James A. 1971. “An Exploration in the Theory of Optimum Income Taxation.” *Review of Economic Studies* 38 (2):175–208.

———. 1976. “Optimal Tax Theory: A Synthesis.” *Journal of Public Economics* 6 (4):327–358.

Ordover, Janusz A. and Edmund S. Phelps. 1979. “The Concept of Optimal Taxation in the Overlapping–Generations Model of Capital and Wealth.” *Journal of Public Economics* 12 (1):1–26.

Pigou, Arthur C. 1928. “An Analysis of Supply.” *Economic Journal* 38 (150):238–257.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Cambridge, MA.: Harvard University Press.

Reis, Catarina. 2010. “Entrepreneurial Labor and Capital Taxation.” *Macroeconomic Dynamics* 15 (3):326–335.

Renes, Sander and Floris T. Zoutman. 2014. “When a Price is Enough: Implementation in Optimal Tax Design.” Discussion Paper TI 14-121/VII, Tinbergen Institute, Amsterdam, the Netherlands.

Saez, Emmanuel. 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies* 68 (1):205–229.

———. 2002. “The Desirability of Commodity Taxation Under Non-linear Income Taxation and Heterogeneous Tastes.” *Journal of Public Economics* 83 (2):217–230.

Saez, Emmanuel and Stefanie Stantcheva. 2018. “A Simpler Theory of Optimal Capital Taxation.” *Journal of Public Economics* 162:120–142.

Saez, Emmanuel and Gabriel Zucman. 2016. “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data.” *Quarterly Journal of Economics* 131 (2):519–578.

Scheuer, Florian and Iván Werning. 2017. “The Taxation of Superstars.” *Quarterly Journal of Economics* 132 (1):211–270.

Schindler, Dirk. 2008. “Taxing Risky Capital Income - A Commodity Taxation Approach.” *FinanzArchiv* 64 (3):311–333.

Seim, David. 2017. "Behavioral Responses to Wealth Taxes: Evidence from Sweden." *American Economic Journal: Economic Policy* 9 (4):395–421.

Spiritus, Kevin and Robin Boadway. 2017. "The Optimal Taxation of Risky Capital Income: The Rate of Return Allowance." Working Paper 6297, CESifo, Munich, Germany.

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics* 17 (2):213–240.

———. 1985. "Inequality and Capital Taxation." Technical Report 457, Stanford University Institute for Mathematical Studies in the Social Sciences.

———. 2000. *Economics of the Public Sector – Third Edition*. New York/London: W.W. Norton & Company.

———. 2018. "Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution." *Journal of Public Economics* 162:101–119.

U.S. Census Bureau. 2019. "Current Population Census, Annual Social and Economic Supplement." Retrieved from https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-10.html.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101 (2):449–472.

Varian, Hal R. 1980. "Redistributive Taxation as Social Insurance." *Journal of Public Economics* 14 (1):49–68.

Vermeulen, Philip. 2018. "How Fat is the Top Tail of the Wealth Distribution?" *Review of Income and Wealth* 64 (2):357–387.

Von Gaudecker, Hans-Martin. 2015. "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?" *Journal of Finance* 70 (2):489–507.

Werning, Ivàn. 2007. "Optimal Fiscal Policy with Redistribution." *Quarterly Journal of Economics* 122 (3):925–967.

Yitzhaki, Shlomo. 1987. "The Relation Between Return and Income." *Quarterly Journal of Economics* 102 (1):77–95.

Zoutman, Floris T. 2018. "The Effect of Capital Income Taxation on Household Savings." Mimeo, Norwegian School of Economics.

A Proof of Lemma 1

We optimize the tax functions T and τ to maximize social welfare using tax perturbations. To solve this problem, we follow the Euler-Lagrange formalism, which was first applied to optimal taxation by Boháček and Kejak (2018), Golosov, Tsyvinski, and Werquin (2014), and Lehmann et al. (2018). We start from a standard proof for the Euler-Lagrange equation, see e.g., (Arfken and Weber, 2005, ch. 17). The standard proof assumes that the arguments of the functions that are being optimized are exogenous. A complication in our case is that a change in the tax functions also affects the taxable incomes z and y. The function arguments thus depend on the functions that we are optimizing. We adapt the standard proof of the Euler-Lagrange equation to incorporate behavioral responses to tax reforms. We only prove the optimality for the tax on capital income $\tau(y)$. The proof for the tax on labor income $T(z)$ follows exactly the same steps.

We introduce the tax perturbations in Appendix A.1. In Appendix A.2, we study the behavioral responses to the different tax perturbations. We then prove Lemma 1 using the Euler-Lagrange approach in Appendix A.3, using the results from the preceding subsections.

A.1 Tax reforms

Tax function $\tau(y)$ is optimal if any small perturbation of τ leaves social welfare unchanged. For any level of capital income y, we introduce a tax reform of size $\epsilon\eta(y)$. The function η is an arbitrary, non-linear, but smooth tax reform function. The parameter ϵ is infinitesimal and allows us to vary the size of the reform. We can construct any small perturbation to τ by choosing η and ϵ. The tax liability at capital income y after the tax reform becomes: $\tau(y) + \epsilon\eta(y)$. If the value of ϵ is zero, then the unperturbed tax function τ is in place. If the optimal value of ϵ is zero for every function η, then the tax schedule $\tau(y)$ is optimal.

In addition to the non-linear perturbations of size $\epsilon\eta(y)$, which we will use to characterize the optimal tax schedules, we also need to introduce scalar perturbations. The reason is that we express our characterizations of the optimal tax schedules in terms of sufficient statistics. The elasticities in Lemma 1 are defined as responses to scalar perturbations of size σ^T, ρ^T, σ^τ and ρ^τ. To prove the optimality of Lemma 1, we thus need to find relations between the behavioral responses to general perturbations of size $\epsilon\eta$ and scalar perturbations of size σ^T, ρ^T, σ^τ and ρ^τ. To do so, we reformulate the individual optimization problem taking into account both the general perturbation function and the scalar perturbation parameters. We then derive comparative statics for the individual choices.

We first rewrite the budget constraints in terms of the taxable incomes, ability, and the tax perturbations. For given ability n, there exists a one-to-one relationship between assets a and capital income $y(a, n)$. Let $a(y, n)$ be the required level of savings for a type-n individual to get capital income y. Then, the individual budget constraints are given by:

$$\mathcal{C}_1(z, y, n, \sigma^T, \rho^T) \equiv z - a(y, n) - \tilde{T}(z) - z\sigma^T - \rho^T, \tag{37}$$

$$\mathcal{C}_2(y, n, \sigma^\tau, \rho^\tau, \epsilon) \equiv a(y, n) + y - \tilde{\tau}(y) - \epsilon\eta(y) - y\sigma^\tau - \rho^\tau. \tag{38}$$

Substitute these budget constraints into the individual utility function to find the reduced-form utility function:

$$\mathcal{U}(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon) \equiv u(\mathcal{C}_1(z, y, n, \sigma^T, \rho^T), \mathcal{C}_2(y, n, \sigma^\tau, \rho^\tau, \epsilon)) - v(z/n). \tag{39}$$

Individuals choose labor income z and capital income y to maximize utility (39) subject to budget constraints (37) and (38). The first-order conditions are:

$$\mathcal{U}_z = \mathcal{U}_y = 0. \tag{40}$$

The functions $\mathcal{U}_z(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon)$ and $\mathcal{U}_y(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon)$ correspond to the shift functions introduced by Jacquet, Lehmann, and Van der Linden (2013).

We denote supply functions for capital and labor income for a type-n individual for given values of the perturbation parameters as $\tilde{y}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$ and $\tilde{z}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$. We denote the corresponding indirect utility function as $\mathcal{V}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$. We apply the envelope theorem to individual objective (39) to find the following property:

$$\mathcal{V}_\epsilon = -u_2 \eta(y). \tag{41}$$

A.2 Behavioral responses to tax reforms

Suppose that there is a tax change. How do individuals change their behavior? To answer this question, suppose there is a marginal change in any parameter $\nu \in \{n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon\}$. Individuals adjust their behavior such that their first-order conditions remain satisfied:

$$0 = \frac{\mathrm{d}\mathcal{U}_z}{\mathrm{d}\nu} = \mathcal{U}_{zz}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} + \mathcal{U}_{zy}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} + \mathcal{U}_{z\nu}, \tag{42}$$

$$0 = \frac{\mathrm{d}\mathcal{U}_y}{\mathrm{d}\nu} = \mathcal{U}_{yz}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} + \mathcal{U}_{yy}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} + \mathcal{U}_{y\nu}. \tag{43}$$

The terms $\mathrm{d}\tilde{z}/\mathrm{d}\nu$ and $\mathrm{d}\tilde{y}/\mathrm{d}\nu$ capture the total effects of a marginal change in the parameter ν on taxable incomes. The terms $\mathrm{d}\tilde{z}/\mathrm{d}\nu$ and $\mathrm{d}\tilde{y}/\mathrm{d}\nu$ include second-round effects caused by the non-linearity of the tax schedules. If taxable incomes change due to a reform, individuals face new marginal tax rates. These changes in the marginal tax rates trigger additional behavioral responses.[29] If we write eqs. (42)–(43) in matrix notation, we can derive the following Lemma.

Lemma 2. *The comparative statics of a marginal change in any parameter ν on labor and capital incomes are given by:*

$$\begin{pmatrix} \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} \\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} \end{pmatrix} = -\begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{U}_{z\nu} \\ \mathcal{U}_{y\nu} \end{pmatrix}. \tag{44}$$

Lemma 2 has the advantage that it reduces the task of finding relations between the effects of different perturbations on the tax bases, to the task of finding relations between the corresponding partial derivatives of $\mathcal{U}_z$ and $\mathcal{U}_y$. We use this result to establish the following Lemma, which shows how any capital income y, the effects of any tax perturbation $\epsilon\eta(y)$ can be decomposed into income effects and substitution effects.

Lemma 3. *The behavioral responses to tax perturbations $\epsilon\eta$ can be decomposed into income and sub-*

[29] Jacquet, Lehmann, and Van der Linden (2013) and Jacobs and Boadway (2014) include similar second-round effects to define behavioral elasticities.

stitution effects:

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\epsilon} = \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\eta(y) + \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right)\eta'(y), \tag{45}$$

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\epsilon} = \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\eta(y) + \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right)\eta'(y). \tag{46}$$

Proof. The function $\mathcal{U}$ has the following second-order partial derivatives which are evaluated at the non-reformed allocation, i.e., $\varepsilon = \rho^T = \rho^\tau = \sigma^T = \sigma^\tau = 0$:

$$\mathcal{U}_{z\rho^T} = -\frac{v' u_{11}}{n u_1}, \qquad \mathcal{U}_{y\rho^T} = \left(u_{11} - u_{12}\frac{u_1}{u_2}\right)\frac{1}{y_a}, \tag{47}$$

$$\mathcal{U}_{z\rho^\tau} = -\frac{v' u_{12}}{n u_1}, \qquad \mathcal{U}_{y\rho^\tau} = \left(u_{21} - u_{22}\frac{u_1}{u_2}\right)\frac{1}{y_a}, \tag{48}$$

$$\mathcal{U}_{z\sigma^T} = z\mathcal{U}_{z\rho^T} - u_1, \qquad \mathcal{U}_{z\sigma^\tau} = y\mathcal{U}_{z\rho^\tau}, \tag{49}$$

$$\mathcal{U}_{y\sigma^\tau} = y\mathcal{U}_{y\rho^\tau} - u_2, \qquad \mathcal{U}_{y\sigma^T} = z\mathcal{U}_{y\rho^T}, \tag{50}$$

$$\mathcal{U}_{z\epsilon} = \mathcal{U}_{z\rho^\tau}\eta(y), \qquad \mathcal{U}_{y\epsilon} = \mathcal{U}_{y\rho^\tau}\eta(y) - u_2\eta'(y). \tag{51}$$

$$\mathcal{U}_{zz} = -u_1 T'' + u_{11}\left(\frac{v'}{n u_1}\right)^2 - \frac{v''}{n^2}, \tag{52}$$

$$\mathcal{U}_{zy} = -\left(u_{11} - u_{12}\frac{u_1}{u_2}\right)\frac{1}{y_a}\frac{v'}{n u_1}, \tag{53}$$

$$\mathcal{U}_{yy} = \left[u_{11} - 2u_{12}\frac{u_1}{u_2} + u_{22}\left(\frac{u_1}{u_2}\right)^2 - (u_2 - u_1)\frac{y_{aa}}{y_a}\right]\frac{1}{y_a^2} - u_2\tau'', \tag{54}$$

$$\mathcal{U}_{zn} = \frac{v'}{n}\left[\frac{u_{11} - u_{21}}{u_1}\frac{y_n}{y_a} + \frac{1}{n}\left(1 + \frac{v'' l}{v'}\right)\right], \tag{55}$$

$$\mathcal{U}_{yn} = -\frac{y_n}{y_a^2}\left[(u_{11} - u_{12}) - (u_{21} - u_{22})\frac{u_1}{u_2} + (u_2 - u_1)\left(\frac{y_{an}}{y_n} - \frac{y_{aa}}{y_a}\right)\right]. \tag{56}$$

Verify the following relations between equations (48), (49), (50) and (51):

$$\mathcal{U}_{z\epsilon} = \mathcal{U}_{z\rho^\tau}\eta(y) + (\mathcal{U}_{z\sigma^\tau} - y\mathcal{U}_{z\rho^\tau})\eta'(y), \tag{57}$$

$$\mathcal{U}_{y\epsilon} = \mathcal{U}_{y\rho^\tau}\eta(y) + (\mathcal{U}_{y\sigma^\tau} - y\mathcal{U}_{y\rho^\tau})\eta'(y). \tag{58}$$

Substitute Lemma 2 for the partial derivatives of $\mathcal{U}$ to find equations (45) and (46). □

Lemma 3 shows that a perturbation of parameter ϵ has two effects. First, the tax liability at each capital income y increases by $\eta(y)$, which causes income effects on individual behavior. Second, the marginal tax rate at each capital income y increases by $\eta'(y)$, which causes substitution effects on individual behavior.

A.3 Proof of Lemma 1: Euler-Lagrange formalism

Consider the optimization problem of a government choosing the optimal value of ϵ, for a given reform function η. The Lagrangian for this maximization problem is given by:

$$\begin{aligned}\Lambda(\epsilon) \equiv & \int_0^{\overline{n}} W(\mathcal{V}(\epsilon, n)) f(n)\mathrm{d}n + \lambda_1 \int_0^{\overline{n}} [\tilde{T}(\tilde{z}(\epsilon, n)) - g_1] f(n)\mathrm{d}n \\ & + \lambda_2 \int_0^{\overline{n}} [\tilde{\tau}(\tilde{y}(\epsilon, n)) + \epsilon\eta(\tilde{y}(\epsilon, n)) - g_2] f(n)\mathrm{d}n.\end{aligned} \tag{59}$$

We use short-hand notations for the function arguments, ignoring the other reform parameters. For now, we assume that the skill distribution has a finite upper limit $\overline{n}$. We assume that this objective function is sufficiently smooth, excluding kinks and bunching.

Evaluate the first-order condition for $\epsilon = 0$, using property (41) and Lemma 3:

$$\begin{aligned}\frac{\partial \Lambda(0)}{\partial \epsilon} = 0 \Leftrightarrow 0 = &\int_{y^0}^{\overline{y}} \left[\frac{1}{\lambda_1/\lambda_2} - \frac{W' u_2}{\lambda_1} + T' \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} + \frac{\tau'}{\lambda_1/\lambda_2} \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} \right] \eta(y) g(y) \mathrm{d}y \\ &+ \int_{y^0}^{\overline{y}} \left[T' \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} \right) + \frac{\tau'}{\lambda_1/\lambda_2} \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} \right) \right] \eta'(y) g(y) \mathrm{d}y. \end{aligned} \tag{60}$$

To arrive at eq. (60), we changed the integration variables from abilities to capital incomes. To do so, we used identity $\mathrm{d}F(n) = \mathrm{d}G(y^n) \Leftrightarrow f(n)\mathrm{d}n = g(y^n)\mathrm{d}y^n$. The latter identity follows from the monotonicity of the allocation. Perform partial integration on the second line of eq. (60), and substitute definitions α_2 and E_τ from the main text:

$$0 = \int_{y^0}^{\overline{y}} \left[\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) + \frac{\mathrm{d}E_\tau(y) g(y)}{\mathrm{d}y} \right] \eta(y) \mathrm{d}y + E_\tau(y^0)\eta(y^0)g(y^0) - E_\tau(\overline{y})\eta(\overline{y})g(\overline{y}). \tag{61}$$

Eq. (61) must hold for every reform function η. Suppose first that the term within square brackets differs from zero on some interval. Then, choose η such that it is zero at the endpoints. The last two terms of eq. (61) become zero. Furthermore, let η have the same sign everywhere as the term within square brackets. With this choice of η, it follows that the value of the integral is strictly positive. However, eq. (61) tells us that the value of the integral must be zero for every reform function η that is zero at the end points. It follows by contradiction that the terms between the square brackets must sum to zero for every capital income y, which is an application of the *fundamental theorem of the calculus of variations.* From this follows the Euler-Lagrange equation for the optimal tax on capital income:

$$\forall y : \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) = -\frac{\mathrm{d}}{\mathrm{d}y} [E_\tau(y) g(y)]. \tag{62}$$

The integral in the first term of eq. (61) must thus be zero. Therefore, the last two terms of eq. (61) must also be zero for every reform function η. This yields the corresponding transversality conditions:

$$E_\tau(y^0) g(y^0) = 0 \text{ and } E_\tau(\overline{y}) g(\overline{y}) = 0. \tag{63}$$

Entirely analogous derivations yield the Euler-Lagrange equation for the tax on labor income:

$$\forall z : (1 - \alpha_1) h(z) = -\frac{\mathrm{d}}{\mathrm{d}z} \left[E_T(z) h(z) \right], \tag{64}$$

with transversality conditions:

$$E_T(0) h(0) = 0 \text{ and } E_T(\overline{z}) h(\overline{z}) = 0. \tag{65}$$

Transversality conditions in eqs. (63) and (65) form two systems of equations in the marginal tax rates at the end points. To arrive at Lemma 1, integrate equations (62) and (64), use transversality conditions (63) and (65), and take the limits $\overline{y} \to \infty$ and $\overline{z} \to \infty$.

B Proof of Proposition 1

This Proof shows that we can rewrite the optimality condition in eq. (23) to obtain the optimality condition in eq. (25). For this, we rewrite the excess burden E_T^n, as given in eq. (19), substitute this into eq. (23), and rearrange.

First, recall that equilibrium labor earnings is written as $z^n = \tilde{z}(\rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$. This implies that we can write:

$$\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} = \frac{\partial \tilde{z}}{\partial \sigma^T} - z^n \frac{\partial \tilde{z}}{\partial \rho^T} = -\left(\frac{z^n}{1 - T'(z^n, \rho^T, \sigma^T)}\right) e_z^n, \tag{66}$$

where the last equality follows from the definition of the elasticity e_z^n in eq. (8).

Second, recall that equilibrium savings is written as $a^n = \tilde{a}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$. This function is implicitly given by the following equation, which follows from substituting the budget constraints of eqs. (1)-(2) into the Euler equation (7):

$$\frac{u_2(z^n - T(z^n, \rho^T, \sigma^T) - a^n, a^n + y(a^n, n) - \tau(y(a^n, n), \rho^\tau, \sigma^\tau))}{u_1(z^n - T(z^n, \rho^T, \sigma^T) - a^n, a^n + y(a^n, n) - \tau(y(a^n, n), \rho^\tau, \sigma^\tau))} = \frac{1}{1 + (1 - \tau'(y(a^n, n), \rho^\tau, \sigma^\tau)) y_a(a^n, n)}. \tag{67}$$

Notice that the marginal tax rate on labor income does not enter eq. (67). This implies that a compensated change in the marginal tax rate on labor income only affects savings through labor income, i.e., $\partial \tilde{a}^c / \partial \sigma^T - z^n \partial \tilde{a}^c / \partial \rho^T = 0$. This implies that we can write:

$$\frac{\mathrm{d}a^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}a^n}{\mathrm{d}\rho^T} = \frac{\partial \tilde{a}^c}{\partial z^n} \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right) = -\frac{\partial \tilde{a}^c}{\partial z^n} \left(\frac{z^n}{1 - T'(z^n, \rho^T, \sigma^T)}\right) e_z^n, \tag{68}$$

where the last equality follows from eq. (66).

Third, since $y^n = y(a^n, n)$, we can write:

$$\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}y^n}{\mathrm{d}\rho^T} = y_a \left(\frac{\mathrm{d}a^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}a^n}{\mathrm{d}\rho^T}\right) = -y_a \frac{\partial \tilde{a}^c}{\partial z^n} \left(\frac{z^n}{1 - T'(z^n, \rho^T, \sigma^T)}\right) e_z^n. \tag{69}$$

Fourth, substituting eqs. (66) and (69) into eq. (19) yields:

$$E_T^n = \left(\frac{T'(z^n, \rho^T, \sigma^T)}{1 - T'(z^n, \rho^T, \sigma^T)} + \frac{1}{1 - T'(z^n, \rho^T, \sigma^T)} \frac{\partial \tilde{a}^c}{\partial z^n} \frac{y_a \tau'}{\lambda_1 / \lambda_2}\right) z^n e_z^n. \tag{70}$$

Substituting eq. (70) into the optimality condition in eq. (23) and rearranging yields eq. (25).

C Proof of Propositions 2, 3 and 4

We first prove Proposition 4, and then derive Propositions 2 and 3 as special cases of Proposition 4. Proposition 4 characterizes the optimal marginal tax on capital income in terms of the optimal marginal tax on labor income. To prove Proposition 4, we thus need to combine the optimal tax schedules from Lemma 1 into a single equation. To be able to do so, we derive the Slutsky symmetry between the tax bases in Subsection C.1. It relates effects of capital taxes on labor income to effects of labor taxes on capital income. Next, in Subsection C.2, we rewrite the optimal capital tax in terms of first-period social welfare weights. Finally, in Subsection C.3, we use the results of the preceding steps to prove

Proposition 4.

C.1 Slutsky symmetry

We derive the Slutsky symmetry between the two tax bases in the following Lemma.

Lemma 4. *Cross-prices responses of labor income and capital income comply to the following Slutsky symmetry:*

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} = \frac{1}{R}\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - z\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\right). \tag{71}$$

Proof. We first construct compensated reforms to the marginal tax rates. Denote taxable incomes in the pre-reform situation as $\tilde{z}^i \equiv \tilde{z}(0,0,0,0,0,n)$ and $\tilde{y}^i \equiv \tilde{y}(0,0,0,0,0,n)$. To determine compensated responses to an increase in marginal tax rates, we fix $\rho^T = -\sigma^T\tilde{z}^i$ and $\rho^\tau = -\sigma^\tau\tilde{y}^i$. Denote the indirect utility function in terms of the compensated perturbations as $\mathcal{V}^*(\sigma^T,\sigma^\tau,\epsilon,n) \equiv \mathcal{V}(\sigma^T,\sigma^\tau,-\sigma^T\tilde{z}^i,-\sigma^\tau\tilde{y}^i,\epsilon,n)$. We first show that these reforms are indeed compensated. Apply the envelope theorem to objective function in eq. (39). This yields the following properties:

$$\frac{\partial\mathcal{V}^*}{\partial\sigma^T} = \frac{\partial\mathcal{V}}{\partial\sigma^T} - \frac{\partial\mathcal{V}}{\partial\rho^T}\tilde{z}^i = -[\tilde{z}(n,\sigma^T,\sigma^\tau,-\sigma^T\tilde{z}^i,-\sigma^\tau\tilde{y}^i,0) - \tilde{z}^i]u_1, \tag{72}$$

$$\frac{\partial\mathcal{V}^*}{\partial\sigma^\tau} = \frac{\partial\mathcal{V}}{\partial\sigma^\tau} - \frac{\partial\mathcal{V}}{\mathrm{d}\rho^\tau}\tilde{y}^i = -[\tilde{y}(n,\sigma^T,\sigma^\tau,-\sigma^T\tilde{z}^i,-\sigma^\tau\tilde{y}^i,0) - \tilde{y}^i]u_2. \tag{73}$$

The last two expressions are zero in the situation before any reforms, when $\rho^T = \rho^\tau = \sigma^T = \sigma^\tau = 0$. We thus indeed constructed compensated reforms of the marginal tax rates.

Evaluate the partial derivative of eq. (72) with respect to σ^τ and of eq. (73) with respect to σ^T, both for the situation without reform:

$$\frac{\partial^2\mathcal{V}^*}{\partial\sigma^\tau\partial\sigma^T} = -\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - \tilde{z}^i\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right)u_1, \tag{74}$$

$$\frac{\partial^2\mathcal{V}^*}{\partial\sigma^T\partial\sigma^\tau} = -\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - \tilde{y}^i\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\right)u_2. \tag{75}$$

Young's theorem implies that the second-order derivatives of any function are symmetric. Apply this requirement to eqs. (74) and (75) to find Slutsky symmetry in eq. (71). □

C.2 Optimal tax on capital income in terms of first-period social welfare weights

To find the optimal combination of taxes on labor income and on capital income, we need to combine the optimal tax schedules in Lemma 1. The tax on capital income in eq. (24) is formulated in terms of social welfare weights in the second period, while the tax on labor income in eq. (23) is formulated in terms of social welfare weights in the first period. We will rewrite the optimal capital income tax in terms of first-period social welfare weights. The first step is to rewrite income effects of reforms in the second period to income effects of reforms in the first period.

The effects of changes in tax liabilities depend on the period in which they occur. If net income increases in the first period, individuals increase their savings. If net income increases in the second period instead, individuals decrease their savings. Furthermore, both responses have different effects on the marginal tax on capital income and on the marginal returns to capital since both taxes on capital income and returns to capital are non-linear. The different effects on the marginal tax rates and on the marginal returns to capital cause different second-round compensated responses. Together these differences affect the government's optimal choice between the tax bases. The following Lemma shows the relation between income effects in both periods.

Lemma 5. *(1) The effects of changes to unearned incomes in the first and in the second period are related as follows:*

$$R\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} = \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T} - \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right)\frac{\partial R}{\partial y}, \tag{76}$$

$$R\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} = \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T} + y_a - \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right)\frac{\partial R}{\partial y}, \tag{77}$$

where $R(y,n)$ is defined as in the main text, and $\frac{\partial R}{\partial y} = -\tau'' y_a + (1-\tau')\frac{y_{aa}}{y_a}$.

(2) The social welfare weights in both periods are related as follows:

$$\alpha_2 R = \alpha_1 - \frac{\partial R}{\partial y}E_\tau - \frac{\tau'}{\lambda_1/\lambda_2}y_a. \tag{78}$$

Proof. Use the individual Euler-condition (7) to find $1-\tau' = (u_1/u_2 - 1)/y_a$. Use this result to find:

$$\frac{\partial R}{\partial y} = -\tau'' y_a + \left(\frac{u_1}{u_2} - 1\right)\frac{y_{aa}}{y_a^2}. \tag{79}$$

Use eq. (79) to verify the following relation between equations (47)–(50) and (53)–(54):

$$\begin{pmatrix}\mathcal{U}_{z\rho^T}\\ \mathcal{U}_{y\rho^T}\end{pmatrix} = y_a\begin{pmatrix}\mathcal{U}_{zy}\\ \mathcal{U}_{yy}\end{pmatrix} + \frac{u_1}{u_2}\begin{pmatrix}\mathcal{U}_{z\rho^\tau}\\ \mathcal{U}_{y\rho^\tau}\end{pmatrix} + \frac{\partial R}{\partial y}\begin{pmatrix}\mathcal{U}_{z\sigma^\tau} - z\mathcal{U}_{z\rho^\tau}\\ \mathcal{U}_{y\sigma^\tau} - y\mathcal{U}_{y\rho^\tau}\end{pmatrix}. \tag{80}$$

Substitute Lemma 2 for the second-order derivatives of $\mathcal{U}$:

$$\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}\left[\frac{u_1}{u_2}\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} + \frac{\partial}{\partial y}\left(\frac{u_1}{u_2}\right)\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} - \begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\end{pmatrix}\right] = y_a\begin{pmatrix}\mathcal{U}_{zy}\\ \mathcal{U}_{yy}\end{pmatrix}. \tag{81}$$

Verify that this is equivalent to:

$$\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}\left[\frac{u_1}{u_2}\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} + \frac{\partial}{\partial y}\left(\frac{u_1}{u_2}\right)\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} - \begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\end{pmatrix} - \begin{pmatrix}0\\ y_a\end{pmatrix}\right] = 0. \tag{82}$$

The part between square brackets must be zero. This proves the first part of the Lemma. The second part of the Lemma follows by substituting equations (76) and (77) into definition (21) of α_1. □

We use Lemma 5 to rewrite the optimal capital income tax in eq. (24) in terms of first-period social welfare weights, in Lemma 6.

Lemma 6. *The optimal capital income tax can be written as follows:*

$$RE_\tau g(y^n) = \int_{y^n}^{\overline{y}}\left[1 - \alpha_1 + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} - \left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right)E_\tau\right]g(y)\mathrm{d}y. \tag{83}$$

Proof. (a) The *first fundamental theorem of calculus* and transversality condition (63) yield the following expression:

$$\text{(84)} \qquad RE_\tau g(y^n) = -\int_{y^n}^{\overline{y}} \frac{\mathrm{d}}{\mathrm{d}y}(RE_\tau g(y))\mathrm{d}y = -\int_{y^n}^{\overline{y}} \left[\frac{\mathrm{d}R}{\mathrm{d}y}E_\tau g(y) + R\frac{\mathrm{d}}{\mathrm{d}y}(E_\tau g(y))\right]\mathrm{d}y.$$

(b) Substitute eq. (78) into the optimal capital income tax (24) and use the individuals' Euler equation to find:

$$\text{(85)} \qquad E_\tau g(y^n) = \int_{y^n}^{\overline{y}} \left[\frac{1}{\lambda_1/\lambda_2} - \frac{1}{R}\left(\alpha_1 - \frac{\partial R}{\partial y}E_\tau - \frac{\tau'}{\lambda_1/\lambda_2}y_a\right)\right]g(y)\mathrm{d}y$$

$$\text{(86)} \qquad = \int_{y^n}^{\overline{y}} \left[\frac{1}{R}\left(1 + \frac{R}{\lambda_1/\lambda_2} - 1 - \alpha_1 + \frac{\partial R}{\partial y}E_\tau + \frac{\tau' y_a}{\lambda_1/\lambda_2}\right)\right]g(y)\mathrm{d}y$$

$$\text{(87)} \qquad = \int_{y^n}^{\overline{y}} \frac{1}{R}\left(1 - \alpha_1 + \frac{\partial R}{\partial y}E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right)g(y)\mathrm{d}y.$$

Take the derivatives with respect to y on both sides:

$$\text{(88)} \qquad \frac{\mathrm{d}}{\mathrm{d}y}(E_\tau g(y)) = -\frac{1}{R}\left(1 - \alpha_1 + \frac{\partial R}{\partial y}E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right)g(y).$$

Substitute the latter into (84) to prove the Lemma. □

C.3 Proof of Proposition 4

We now have all the required elements to prove Proposition 4. We start by combining the expressions for the optimal taxes on capital and labor income. Then, we use the properties from the preceding subsections to find Proposition 4.

Proof. Substitute $h(z^n) = g(y^n)(\mathrm{d}y^n/\mathrm{d}n)/(\mathrm{d}z^n/\mathrm{d}n)$, $\mathrm{d}y^n/\mathrm{d}n = (\partial\tilde{y}^c/\partial z)(\mathrm{d}z/\mathrm{d}n) + \partial\tilde{y}^c/\partial n$, the definitions in eqs. (19)–(20) for E_T^n and E_τ, and the optimal labor tax in eq. (23) into the optimal capital tax in eq. (83), and rearrange:

$$\begin{aligned}
&\frac{\tau'}{\lambda_1/\lambda_2}\left(R\left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau}\right) - \left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}\right)\frac{\partial\tilde{y}^c}{\partial z}\right) \\
&\quad = -T'\left(R\left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau}\right) - \left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}\right)\frac{\partial\tilde{y}^c}{\partial z}\right) - \frac{1}{h(z^n)}\int_{z^n}^{\infty}(1-\alpha_1^m)h(z^m)\mathrm{d}z^m\frac{\partial\tilde{y}^c}{\partial n}\left(\frac{\mathrm{d}z^n}{\mathrm{d}n}\right)^{-1} \\
&\qquad - \frac{1}{g(y^n)}\int_{y^n}^{\infty}\frac{1+\tilde{y}_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}g(y)\mathrm{d}y + \frac{1}{g(y^n)}\int_{y^n}^{\infty}\left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right)E_\tau g(y)\mathrm{d}y.
\end{aligned}$$

Substitute Slutsky symmetry from eq. (71), employ separability of preferences between leisure and consumption to substitute $\partial\tilde{y}^c/\partial\sigma^T - z^n\partial\tilde{y}^c/\partial\rho^T = 0$ and substitute Proposition 1:

$$\begin{aligned}
\frac{\tau' y_a^n}{\lambda_1/\lambda_2}\left[\frac{R}{y^n}\frac{1}{y_a^n}\left(\frac{\partial\tilde{y}^c}{\partial\sigma^\tau} - y^n\frac{\partial\tilde{y}^c}{\partial\rho^\tau}\right)\right] &= -\left(\frac{n}{y^n}\frac{\partial\tilde{y}^c}{\partial n}\right)\left(\frac{n}{z^n}\frac{\mathrm{d}z^n}{\mathrm{d}n}\right)^{-1}\left(\frac{T'(z)}{1-T'(z)} + \frac{sy_a\tau'(y)}{\lambda_1/\lambda_2}\right)e_z \\
&\quad - \frac{1}{y^n g(y^n)}\int_{y^n}^{\infty}\frac{1+\tilde{y}_a-\lambda_1/\lambda_2}{\lambda_1/\lambda_2}g(y)\mathrm{d}y + \frac{1}{y^n g(y^n)}\int_{y^n}^{\infty}\left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right)E_\tau g(y)\mathrm{d}y.
\end{aligned}$$

By substituting the optimal labor tax, we thus rewrote the optimal capital tax in terms of capital income elasticities conditional on labor income. Substitute the definitions for $\xi_{y|z}^n$, ξ_z^n, and $e_{y|z}^n$ to find Proposition 4. □

C.4 Proofs of Propositions 2 and 3

We now prove Propositions 2 and 3. Eq. (31) implies that $\partial R(y^n, n)/\partial n = 0$ in both our first microfoundation ($y_{aa} = 0$) and our second microfoundation ($y_n = 0$). Substituting $\partial R(y^n, n)/\partial n = 0$, the bottom transversality conditions of eqs. (63) and (65), and the optimal excess burden of eq. (23) into eq. (83) yields:

$$\lambda_1/\lambda_2 = 1 + \bar{y}_a(y). \tag{89}$$

To prove Proposition 2, substitute eq. (89) and $\partial R(y^n, n)/\partial n = 0$ into eq. (30), noting that marginal returns to capital y_a are equal for all individuals. To prove Proposition 3, substitute eq. (89), $\partial R(y^n, n)/\partial n = 0$ and $\xi_{y|n} = 0$ into eq. (30).

D Rewriting optimal tax formulas: from sufficient statistics to model primitives

In the main text, we have written the optimal tax schedules in terms of sufficient statistics. However, in this section we rewrite the optimal tax schedules in terms of underlying behavior in order to numerically simulate the optimal tax schedules. An important complication in rewriting the optimal tax schedules is that the sufficient statistics depend on the second-order derivatives of the tax functions T and τ.[30] Approximating the second-order derivatives of the tax functions is computationally expensive. Hence, we rewrite the optimal tax schedules in a way that does not include the second derivatives of the tax functions, but only includes the utility function, the social welfare function, the individual abilities and their distribution, the excess returns to saving, and individual behavior.

We start in Subsections D.1 and D.2 by formally introducing a mass of disabled individuals at the bottom to ensure non-zero marginal tax rates at the bottom. Next, in Subsection D.3, we study the relation between the multipliers on the government budget constraint λ_1 and λ_2. Then, in Subsection D.4, we reformulate the optimal tax schedules in terms of model primitives, for a given social marginal value of income redistribution. In Subsection D.5 we reformulate the social marginal value of income redistribution in terms of model primitives. We find the optimal value for the first-period government budget multiplier λ_1 in Subsection D.6.

D.1 Disabled individuals

A fraction D of the population is disabled. We denote variables for a disabled individual using a superscript d. Disabled individuals have zero earning ability ($n = 0$) and do not earn any labor income: $z^d = 0$. In the first period, disabled individuals receive a transfer $-T(0)$, which they consume immediately or save for consumption in the second period. The level of savings of the disabled is equal to the level of savings of the least able individual if the tax system is optimized: $a^d = a^{\underline{n}}$, where $\underline{n}$ is the ability level of the least able individual who is not disabled. We assume that the savings of the disabled are exogenously given so that they do not respond to tax perturbations. There is no reason to believe that this simplifying assumption has any numerically meaningful effect on our results but it does make our simulations considerably more tractable. Neither the disabled, nor the least able individual earn any

[30] Lemma 2 demonstrates that the sufficient statistics depend on the Hessian of the reduced-form utility function $\mathcal{U}$, while the proof of Lemma 3 shows that the elements of this Hessian depend on the second-order derivatives of the tax functions T and τ.

excess returns: $y^d = \underline{y} = ra^{\underline{n}}$, where we simplify notation $\underline{y} \equiv y^{\underline{n}}$. The disabled pay the same tax as the workers with the same capital income $-\tau(y^d)$.

The budget constraints for the disabled individuals are:

$$\mathcal{C}_1^d(y^d, \rho^T) = -\tilde{T}(0) - \rho^T - \frac{y^d}{r}, \tag{90}$$

$$\mathcal{C}_2^d(y^d, \sigma^\tau, \rho^\tau, \epsilon) = \frac{y^d}{r} + y^d - \tilde{\tau}(y^d) - \epsilon\eta(y^d) - y^d\sigma^\tau - \rho^\tau. \tag{91}$$

We denote the indirect utility function for the disabled as $\mathcal{V}^d(\sigma^\tau, \rho^T, \rho^\tau, \epsilon)$. Use the envelope theorem to find the following properties:

$$\mathcal{V}_\epsilon^d = -u_2^d\eta(y^d),\ \mathcal{V}_{\rho^\tau}^d = -u_2^d. \tag{92}$$

D.2 Optimal tax schedules with disabled individuals

The new Lagrangian for the government optimization problem, taking into account the mass of individuals at the bottom, is given by:

$$\begin{aligned}\Lambda(\epsilon) \equiv & DW(\mathcal{V}^d(\epsilon)) + \int_{\underline{n}}^{\overline{n}} W(\mathcal{V}(\epsilon,n))f(n)\mathrm{d}n + \lambda_1[D\tilde{T}(0) - g_1] + \lambda_1\int_{\underline{n}}^{\overline{n}} \tilde{T}(\tilde{z}(\epsilon,n))f(n)\mathrm{d}n \\ & + \lambda_2 D[\tilde{\tau}(y^d) + \epsilon\eta(y^d)] - \lambda_2 g_2 + \lambda_2\int_{\underline{n}}^{\overline{n}} [\tilde{\tau}(\tilde{y}(\epsilon,n)) + \epsilon\eta(\tilde{y}(\epsilon,n))]f(n)\mathrm{d}n.\end{aligned} \tag{93}$$

We again assume that this objective function is sufficiently smooth, excluding kinks and bunching. Evaluate the first-order condition at $\epsilon = 0$, using properties (41) and (92) and Lemma 3:

$$\begin{aligned}\frac{\partial\Lambda(0)}{\partial\epsilon} = 0 \Leftrightarrow 0 = & \left(\frac{1}{\lambda_1/\lambda_2} - \frac{W'u_2^d}{\lambda_1}\right)\eta(\underline{y})D + \int_{\underline{y}}^{\overline{y}}\left[\frac{1}{\lambda_1/\lambda_2} - \frac{W'u_2}{\lambda_1} + T'\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} + \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right]\eta(y)g(y)\mathrm{d}y \\ & + \int_{\underline{y}}^{\overline{y}}\left[T'\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right) + \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right)\right]\eta'(y)g(y)\mathrm{d}y,\end{aligned} \tag{94}$$

where $\overline{y} \equiv y^{\overline{n}}$ is the highest level of capital income. Perform partial integration on the second line of eq. (94) and substitute definitions α_2 and E_τ from the main text:

$$\begin{aligned}0 = & \int_{\underline{y}}^{\overline{y}}\left[\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2\right)g(y) + \frac{\mathrm{d}E_\tau(y)g(y)}{\mathrm{d}y}\right]\eta(y)\mathrm{d}y \\ & + \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^d\right)\eta(\underline{y})D + E_\tau(\underline{y})\eta(\underline{y})g(\underline{y}) - E_\tau(\overline{y})\eta(\overline{y})g(\overline{y}).\end{aligned} \tag{95}$$

Applying the fundamental theorem of the calculus of variations in the same way as we did in Subsection A.3, we find that the Euler-Lagrange Equation (62) for the optimal tax on capital income is not directly affected by the introduction of a mass of individuals at the bottom. Furthermore, the transversality condition (63) at the top remains unaltered. At the bottom though, we find the following transversality condition:

$$E_\tau(\underline{y})g(\underline{y}) = -\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^d\right)D. \tag{96}$$

Entirely analogous derivations show that the Euler-Lagrange equation (64) for the tax on labor income and the transversality condition (64) at the top remain unaltered with the addition of the disabled individuals. At the bottom, we find the following transversality condition for the tax on labor income:

$$E_T(0)h(0) = -(1-\alpha_1^d)D. \tag{97}$$

Given that the Euler-Lagrange equations and the transversality conditions at the top are not directly affected by the presence of a mass of individuals at the bottom, the optimal tax schedules in Lemma 1 remain valid. Transversality conditions (96) and (97) form a new system of equations in the marginal tax rates at the bottom. The optimal marginal tax rates at the bottom are now positive, as we demonstrate in subsection D.4.

D.3 Relation between the government budget multipliers

For the simulations we need explicit expressions for the multipliers on the government budget constraints λ_1 and λ_2. In this section, we study the relation between the two multipliers. Substitute Lemmas 5 and 6 into the transversality condition for the lowest ability level in eq. (96), use our assumption that the disabled do not respond to tax perturbations, and use $\mathrm{d}R/\mathrm{d}y = \partial R/\partial y$ in the case with closely-held assets:

$$\int_{\underline{n}}^{\overline{n}}\left(1-\alpha_1+\frac{1+y_a-\lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right)f(n)\mathrm{d}n = \left(\alpha_1^d-\frac{1+y_a^d}{\lambda_1/\lambda_2}\right)D. \tag{98}$$

Substitute transversality condition for the lowest ability in eq. (97):

$$\int_{\underline{n}}^{\overline{n}}\left(\frac{1+y_a}{\lambda_1/\lambda_2}-1\right)f(n)\mathrm{d}n = \left(1-\frac{1+y_a^d}{\lambda_1/\lambda_2}\right)D. \tag{99}$$

In the case with closely-held assets, our assumptions imply that all individuals earn the same marginal return to capital y_a. It follows that in the optimum the multipliers on the government budget constraint are related as follows:

$$\lambda_1/\lambda_2 = 1+y_a. \tag{100}$$

In the optimum, the marginal rate of substitution between the two periods for the government should thus be equal to the private marginal rate of transformation. Substituting eq. (100) into Lemma 6, we then find that in the case of a closely-held asset, a simple relation exists between the net social marginal benefits of redistribution in both periods:

$$\int_{n}^{\overline{n}}\left(\frac{1}{\lambda_1/\lambda_2}-\alpha_2\right)f(n^*)\mathrm{d}n^* = \frac{u_2}{u_1}\int_{n}^{\overline{n}}(1-\alpha_1)f(n^*)\mathrm{d}n^*. \tag{101}$$

D.4 Tax schedules for given benefits of redistribution

We take two steps to rewrite the optimal marginal tax schedules in terms of model primitives. In this Subsection, we take the net social marginal benefit of a higher marginal tax rate at n, $\int_n^{\overline{n}}(1-\alpha_1)f(n^*)\mathrm{d}n^*$, as given. In the next Subsection we also derive the net social marginal benefits of a higher tax rate in terms of model primitives.

Rewrite Lemma 1 in matrix notation, taking into account Lemma 6, using the assumption y_{aa} =

$y_{an} = 0$, and using eq. (100):

$$\begin{pmatrix} \frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} & \frac{\mathrm{d}z^n}{\mathrm{d}\sigma^\tau} - y^n \frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau} \\ \frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}y^n}{\mathrm{d}\rho^T} & \frac{\mathrm{d}y^n}{\mathrm{d}\sigma^\tau} - y^n \frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau} \end{pmatrix}^{tr} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = -\begin{pmatrix} 1 & 0 \\ 0 & \frac{u_2}{u_1} \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}z}{\mathrm{d}n} \\ \frac{\mathrm{d}y}{\mathrm{d}n} \end{pmatrix} \frac{\int_n^{\overline{n}} (1-\alpha_1) f(n^*) \mathrm{d}n^*}{f(n)}, \tag{102}$$

where superscript tr stands for matrix transpose. Substitute Lemma 2:

$$\begin{aligned} &\begin{pmatrix} \mathcal{U}_{z\sigma^T} - z^n \mathcal{U}_{z\rho^T} & \mathcal{U}_{y\sigma^T} - z^n \mathcal{U}_{y\rho^T} \\ \mathcal{U}_{z\sigma^\tau} - y^n \mathcal{U}_{z\rho^\tau} & \mathcal{U}_{y\sigma^\tau} - y^n \mathcal{U}_{y\rho^\tau} \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} \\ &= -\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{R} \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{U}_{zn} \\ \mathcal{U}_{yn} \end{pmatrix} \frac{\int_n^{\overline{n}} (1-\alpha_1) f(n^*) \mathrm{d}n^*}{f(n)}. \end{aligned} \tag{103}$$

Substitute equations (49)–(50):

$$\begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} \frac{\int_n^{\overline{n}} (1-\alpha_1) f(n^*) \mathrm{d}n^*}{f(n)}. \tag{104}$$

Eliminate the two left-most matrices on both sides to find the following characterization of the optimal tax schedules:

$$\begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = \begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} \frac{\int_n^{\overline{n}} (1-\alpha_1) f(n^*) \mathrm{d}n^*}{f(n)}. \tag{105}$$

Eq. (105) is very similar to the optimal tax expressions in Mirrlees (1976). The main difference is that Mirrlees (1976) only accounts for the dependence of the marginal rates of substitution on ability, capturing e.g., varying degrees of complementarity between goods and leisure (Corlett and Hague, 1953), while equation (105) also accounts for the direct dependence of the individual budget constraints in eqs. (37)–(38) on ability.

We can now derive the optimal marginal tax rates for the individuals with the lowest ability. Substitute the transversality condition for the lowest ability from eq. (97) into eq. (105) to find :

$$\begin{pmatrix} T'(0) \\ \frac{\tau'(y)}{1+y_a} \end{pmatrix} = -\begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} (1-\alpha_1^d) \frac{D}{f(0)}. \tag{106}$$

Marginal tax rates at the bottom are thus positive if single crossing conditions $\mathcal{U}_{zn}, \mathcal{U}_{yn} > 0$ are met, and social welfare weights α_1^d for the disabled are larger than one. Substitute eq. (100) into eq. (98) to find $1 = \alpha_1^d D + \int_0^{\overline{n}} \alpha_1 f(n) \mathrm{d}n$. If we assume that the welfare weights of the disabled are higher than the welfare weights of the able, then it follows that $\alpha_1^d > 1$, and thus marginal tax rates at the bottom are positive. Positive marginal tax rates at the bottom enable redistribution from the able to the disabled. If the proportion of disabled individuals equals zero, $D = 0$, the marginal tax rate at the bottom is zero.

D.5 Social marginal benefit of income redistribution

Eq. (105) is not completely written in terms of model primitives. The social marginal value of a higher tax rate still contains sufficient statistics:

$$\int_n^{\overline{n}} (1-\alpha_1) f(\tilde{n}) \mathrm{d}\tilde{n} = \int_n^{\overline{n}} \left(1 - \frac{W' u_1}{\lambda_1} + T' \frac{\mathrm{d}z^{n^*}}{\mathrm{d}\rho^T} + \frac{\tau'}{\lambda_1/\lambda_2} \frac{\mathrm{d}z^{n^*}}{\mathrm{d}\rho^T} \right) f(n^*) \mathrm{d}n^*. \tag{107}$$

To find an alternative formulation for eq. (107), we rewrite it as a differential equation. Introduce the following composite function:

$$\Phi(n) \equiv -\frac{1}{u_1}\int_n^{\overline{n}}(1-\alpha_1)f(n^*)\mathrm{d}n^*, \tag{108}$$

with derivative:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}n} = -\frac{\mathrm{d}u_1}{\mathrm{d}n}\frac{\Phi}{u_1} + \frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}+T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}+\frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}\right)f(n). \tag{109}$$

Substitute eq. (105) to rewrite the terms with the income effects, and use definition (39) to expand $\mathrm{d}u_1/\mathrm{d}n$:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}n} = \left(\mathcal{U}_{z\rho^T}\frac{\mathrm{d}z^n}{\mathrm{d}n}+\mathcal{U}_{y\rho^T}\frac{\mathrm{d}y^n}{\mathrm{d}n}+u_{11}\frac{y_n}{r}\right)\frac{\Phi}{u_1}+\frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}\right)f(n)-\frac{\Phi}{u_1}\left(\mathcal{U}_{zn}\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}+\mathcal{U}_{yn}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}\right). \tag{110}$$

Eq. (110) is a differential equation with Φ as unknown function. To further simplify this differential equation, we first formulate a Lemma that relates the effects of perturbations of any two parameters μ and ν on the tax bases.

Lemma 7. *For any two perturbation parameters μ and ν, the effects on the taxable incomes are related as follows:*

$$\mathcal{U}_{z\mu}\frac{\mathrm{d}z^n}{\mathrm{d}\nu}+\mathcal{U}_{y\mu}\frac{\mathrm{d}y^n}{\mathrm{d}\nu}=\mathcal{U}_{z\nu}\frac{\mathrm{d}z^n}{\mathrm{d}\mu}+\mathcal{U}_{y\nu}\frac{\mathrm{d}y^n}{\mathrm{d}\mu}.$$

Proof. Use Lemma 2 to find the following relation:

$$\begin{aligned}\mathcal{U}_{z\mu}\frac{\mathrm{d}z^n}{\mathrm{d}\nu}+\mathcal{U}_{y\mu}\frac{\mathrm{d}y^n}{\mathrm{d}\nu} &= -\begin{pmatrix}\mathcal{U}_{z\mu} & \mathcal{U}_{y\mu}\end{pmatrix}\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}^{-1}\begin{pmatrix}\mathcal{U}_{z\nu}\\ \mathcal{U}_{y\nu}\end{pmatrix}\\ &=\left[-\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz}\mathcal{U}_{yy} & \end{pmatrix}^{-1}\begin{pmatrix}\mathcal{U}_{z\mu}\\ \mathcal{U}_{y\mu}\end{pmatrix}\right]^{tr}\begin{pmatrix}\mathcal{U}_{z\nu}\\ \mathcal{U}_{y\nu}\end{pmatrix}=\mathcal{U}_{z\nu}\frac{\mathrm{d}z^n}{\mathrm{d}\mu}+\mathcal{U}_{y\nu}\frac{\mathrm{d}y^n}{\mathrm{d}\mu}.\end{aligned} \tag{111}$$

□

Lemma 7 allows us to simplify differential equation (110):

$$\frac{\mathrm{d}\Phi(n)}{\mathrm{d}n} = -\frac{u_{11}}{u_1}\frac{y_n}{r}\Phi(n)+\frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}\right)f(n). \tag{112}$$

This differential equation has a standard form, with solution:

$$\int_n^{\overline{n}}(1-\alpha_1)f(n^*)\mathrm{d}n^* = u_1\int_n^{\overline{n}}\frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}\right)\exp\left[\int_n^{n^*}\frac{u_{11}}{u_1}\frac{y_{n^{**}}}{r}\mathrm{d}n^{**}\right]f\left(n^*\right)\mathrm{d}n^*. \tag{113}$$

We have thus rewritten the social marginal benefits of a higher marginal tax rate in terms of economic fundamentals. Substituting eq. (113) into eq. (105), substituting Eqs. (55)–(56) and using the assumption $u_{12}=0$, we thus find the optimal tax schedules in terms of model primitives:

$$\begin{pmatrix}T'\\ \frac{\tau'}{1+y_a}\end{pmatrix}=\begin{pmatrix}\frac{v'}{n}\left(\frac{u_{11}}{u_1}\frac{y_n}{r}+\frac{1}{n}\left(1+\frac{v''l}{v'}\right)\right)\\ -\frac{y_n}{r^2}\left(u_{11}+\frac{u_1}{u_2}u_{22}\right)\end{pmatrix}\frac{\int_n^{\overline{n}}\frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}\right)\exp\left[\int_n^{n^*}\frac{u_{11}}{u_1}\frac{y_{n^{**}}}{r}\mathrm{d}n^{**}\right]f\left(n^*\right)\mathrm{d}n^*}{f(n)}. \tag{114}$$

D.6 Value of the first-period multiplier on the government budget constraint

Eq. (114) characterizes the optimal tax schedules almost entirely in terms of model primitives. The only remaining variable to rewrite is the first-period multiplier on the government budget constraint. Substitute the first-period benefit of redistribution eq. (113) into transversality condition eq. (97) for the least able individual:

$$\int_{\underline{n}}^{\overline{n}} \frac{1}{u_1}\left(1-\frac{W'u_1}{\lambda_1}\right)\exp\left[\int_{\underline{n}}^{n}\frac{u_{11}}{u_1}\frac{y_{n^*}}{r}\mathrm{d}n^*\right]f(n)\,\mathrm{d}n = -\frac{1}{u_1^{\underline{n}}}\left(1-\frac{W'^d u_1^d}{\lambda_1}\right)D. \tag{115}$$

Solve this for λ_1:

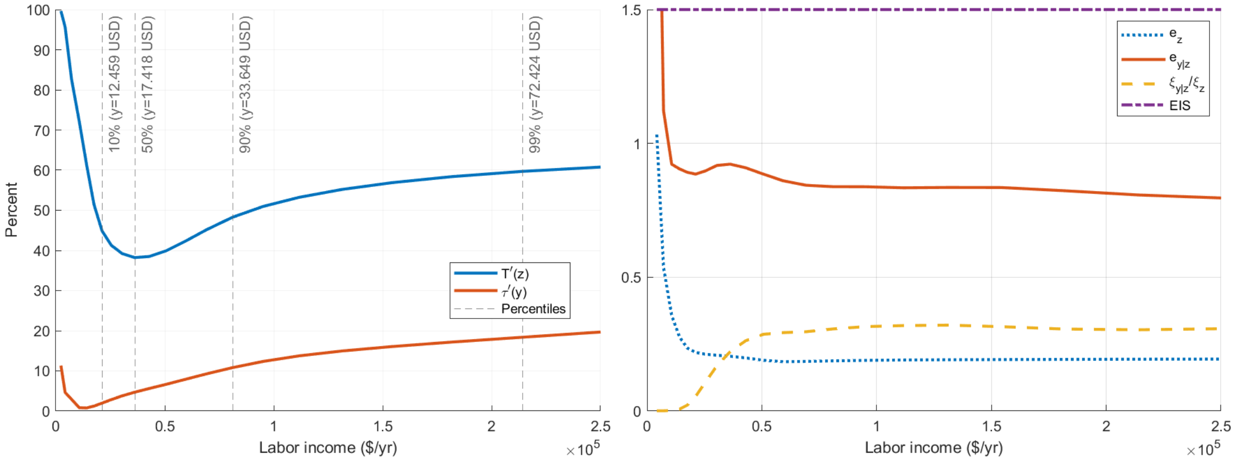
$$\lambda_1 = \frac{\int_{\underline{n}}^{\overline{n}} W'\exp\left[\int_{\underline{n}}^{n}\frac{u_{11}}{u_1}\frac{y_{n^*}}{r}\mathrm{d}n^*\right]f(n)\,\mathrm{d}n + W'^d\frac{u_1^d}{u_1^{\underline{n}}}D.}{\int_{\underline{n}}^{\overline{n}} \frac{1}{u_1}\exp\left[\int_{\underline{n}}^{n}\frac{u_{11}}{u_1}\frac{y_{n^*}}{r}\mathrm{d}n^*\right]f(n)\,\mathrm{d}n + \frac{1}{u_1^{\underline{n}}}} \tag{116}$$

The larger is the fraction of disabled individuals D, the larger will be the multiplier λ_1, and the lower will be welfare weights α_1 of the able individuals. Therefore, the larger is the proportion of disabled individuals, the higher will be the marginal tax rates for the rest of the population. In the main text, we explain how we simulate optimal tax schedules by applying a fixed point algorithm to eqs. (114) and (116).

E Additional simulations

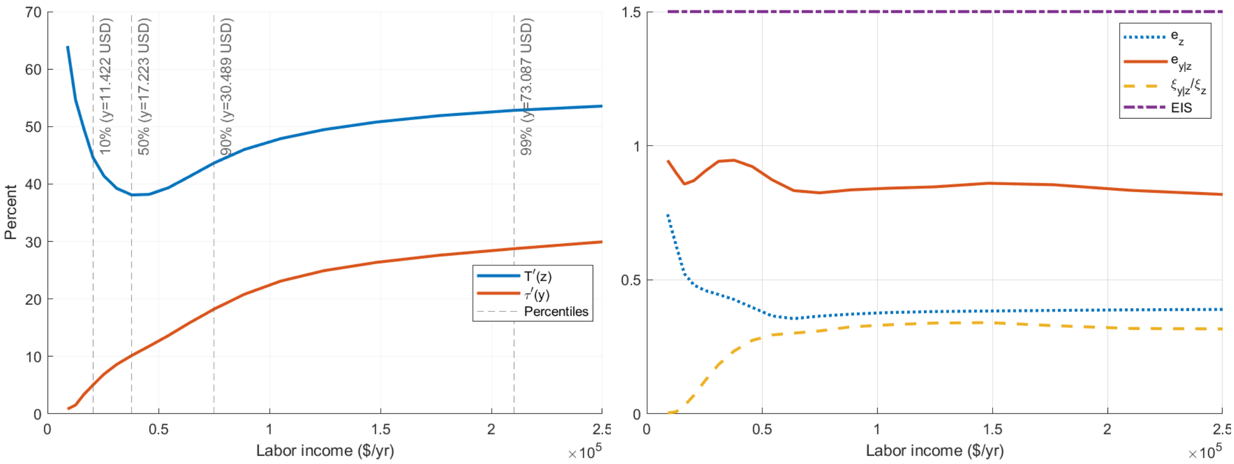
Figure 4 presents optimal tax schedules and elasticities if the government objective is utilitarian. Figure 5 presents results if the Frisch elasticity of labor supply equals $\varepsilon = 0.5$. All other parameters are the same as in our baseline calibration. We find that optimal marginal tax rates decline if we reduce the inequality aversion and if we increase the Frisch elasticity. Our main conclusions remain valid: optimal marginal taxes on capital income are substantial and they increase with income.

Figure 4: Results with utilitarian government ($\gamma = 0$) in the baseline (high EIS, full excess returns)



(a) Optimal marginal capital and labor income taxes (b) Elasticities

Figure 5: Results with high Frisch elasticity ($\varepsilon = 0.5$) in the baseline (high EIS, full excess returns)



(a) Optimal marginal capital and labor income taxes (b) Elasticities