Optimal Taxation of Capital Income with Heterogeneous Rates of Return

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

January 30, 2019

Very preliminary: please do not quote or post files online

Abstract

We derive optimal taxes on labor and capital income if individuals have heterogeneous returns on capital, since there is increasing evidence that people differ in the rates of return on their savings, even after controlling for risk. We allow for two distinct reasons why returns are heterogeneous: because individuals higher ability obtain higher returns on their savings and because wealthier individuals achieve higher returns due to scale effects in wealth management. In either case, a strictly positive tax on capital income is part of a Pareto-efficient redistributive tax system.

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 about how governments should tax capital back to the center of the policy debate. Arguments against capital income taxation dates 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. Hence, if all inequality in capital income derives only from inequality in labor income, then it is of no independent concern for setting taxes on capital income. Since taxes on capital income would be distributionally equivalent to taxes on labor income, they would distort labor supply in the same way, but, in addition, would distort saving decisions as well. Hence, it would be better not to tax capital income at all (Atkinson and Stiglitz, 1976).¹

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 an important assumption: all individuals obtain the same rate of return on their savings, regardless of their earning

*All authors: Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. 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 rusu@ese.eur.nl. Spiritus, E-mail kevin.spiritus@kuleuven.be, Internet: http://www.kevinspiritus.eu. We thank Katherine Cuff, André Decoster, Tilman Klumpp, Claus Kreiner, Agnar Sandmo, Floris Zoutman and participants of numerous seminars and conferences for helpful comments and discussions. Kevin Spiritus acknowledges the financial support of the Belgian Federal Science Policy Office (BELSPO) via the BRAIN.be project BR/121/A5/CRESUS.

¹The 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 income, all individuals then save the same amount (Atkinson and Stiglitz, 1976).
abilities or wealth. This assumption has become untenable. A large and growing body of evidence shows that people differ in their returns on savings, and that these returns are systematically related to measures of ability and wealth, and are unrelated to risk-taking behavior. This evidence strongly suggests that inequality in capital income does not simply derive from inequality in labor income.

First, there is growing direct evidence that rates of return to saving indeed differ across individuals. A seminal contribution is Yitzhaki (1987), who finds that rates of return increase with income. He is unable to determine whether this heterogeneity is simply due to differences in risk-taking or better investment possibilities of richer individuals. Fagereng et al. (2016) present strong evidence 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 portfolio choices. They analyze administrative micro data containing the records of the entire population of Norwegian tax payers over twenty years. Rates of return to capital increase with wealth by more than two percentage points from the fiftieth to the ninety-fifth wealth percentile. 81 percent of this increase can be attributed to individual-fixed effects. Furthermore, 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.

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 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. 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 lifecycle models with observed patterns of wealth inequality (e.g., Benhabib, Bisin, and Luo, 2015; Gabaix et al., 2015; Lusardi, Michaud, and Mitchell, 2017; Kacperczyk, Nosal, and Stevens, 2018). In particular, Lusardi, Michaud, and Mitchell (2017) suggest that a staggering 30-40 percent of wealth inequality can be explained by return heterogeneity.

We derive the implications of return heterogeneity for optimal 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. Ability and labor supply determine total labor income. Ability and savings determine total capital income. The government cannot observe ability, only labor and capital income. As a result, it must rely on distortionary taxes on labor and capital income to optimally redistribute income. Individuals supply labor and decide how much to save in

\[ \text{Continued...} \]

---

2The underlying determinants of these persistent individual-specific factors remain unclear. They may be due to different investments in financial sophistication (e.g., Lusardi and Mitchell, 2014), or due to differences in innate abilities, intelligence, and genes (e.g., Barth, Papageorge, and Thom, 2018).

3Bach, Calvet, and Sodini (2010) study Swedish administrative data and also find that rates of return are, on average, increasing over the wealth distribution. In contrast to Fagereng et al. (2016), they attribute differences in capital returns mostly to differences in risk-taking behavior. However, they also acknowledge that their tests lack statistical power.
the first period of their life-cycle. They consume all their savings in the second period.\(^4\) We abstract from risk-taking to focus our analysis solely on the implications of return heterogeneity for the setting of taxes on capital income.\(^5\) We are agnostic about the process that generates heterogeneous capital returns. We assume that the return on savings is a function of both ability and the level of savings. This specification captures the notion that rates of return might originate from differences in earning ability or from wealth itself, in line with the empirical evidence cited above. Our general specification captures a number of plausible microfoundations for return heterogeneity. For example, high-ability individuals may have better access to investments in closely-held companies that generate excess returns in capital income. Alternatively, wealthier individuals may benefit from scale effects and information advantages by delegating the management of their wealth to private banks and asset managers. We demonstrate that optimal taxes on capital income are positive both if rates of return to savings are increasing in ability and if rates of return to savings are increasing in savings.

First, we consider the case where rates of return to capital are only a function of ability and not of savings. Taxes on labor and capital income then cease to be distributionally equivalent. Intuitively, capital income cannot be perfectly predicted by labor income if capital returns depend on ability: capital incomes differ even if labor incomes were the same. Therefore, capital income contains information about ability in addition to what is revealed through labor income. Consequently, the government should tax both capital income and labor income to optimally redistribute income.\(^6\) The optimal tax on capital income trades off (additional) redistributional gains against distortions in saving. We derive a condition for the Pareto efficient dual income tax structure in terms of sufficient statistics (tax wedges, elasticities of taxable labor and capital income with respect to their net-of-tax rate, and (conditional) elasticities of labor and capital income with respect to ability). The sign and size of the optimal tax on capital are critically determined by the conditional elasticity of capital income with respect to ability. This elasticity captures return heterogeneity, as it measures the extent to which capital income varies with ability, conditional on labor income. If returns positively correlate with ability, the elasticity is positive, so the optimal tax on capital income is positive. If the elasticity is zero, returns to capital do not vary with ability, so optimal capital income taxes are also zero. Thus, our model nests Atkinson and Stiglitz (1976) as a special case.

Second, we consider the case where rates of return to capital are only a function of savings and not of ability. In that case, there are scale effects in savings: the more individuals save, the higher are their returns to saving. Implicitly, this means that there is a failure of the capital market that prevents the poor to invest in assets with high returns. Hence, not all return differences will be arbitraged away. We show that the optimal tax on capital income is positive also in this case. Intuitively, the government alleviates capital market failure by lowering taxes on labor income in the first period of the life-cycle. This is to let the rich save more and exploit their scale effects to obtain high returns to saving. The government then sets positive taxes on capital income in the second period of the life-cycle to recoup these higher returns to saving. In doing so, the government ensures that the poor can also share in the higher returns to saving of the rich. As a result, it can either redistribute more income towards the poor for the same overall distortions of the tax system or lower taxes on labor income for the

\(^4\)The life-cycle structure of our model is similar to Ordover and Phelps (1979) who also analyze 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).

\(^5\)Allowing for risk-taking and portfolio choice would only strengthen our case for positive taxes on capital income, as we argue in the literature review.

\(^6\)Stiglitz (1985, 2000, 2018) has also conjectured but not formally shown that return heterogeneity would lead to optimal taxes on capital income that are positive if they depend on ability. We confirm this conjecture.
same amount of overall income redistribution the tax system aims to achieve. As before, we derive a condition for Pareto efficient taxes on capital income that is expressed in terms of sufficient statistics (the interest elasticity of capital income, the distribution of capital income, and the distribution of rates of return). The optimal tax on capital income trades off distortions in savings against the efficiency gains of alleviating the failure of the capital market. We also derive a formula for the Pareto optimal top rate on capital income, which only depends on the elasticity of capital income, the Pareto parameter of the distribution of capital income, and the social and top rates of return to capital.

The remainder of the paper is organized as follows. In Section 2, we discuss earlier results on the 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 tax on capital income. A final section concludes. The Appendix contains some derivations and proofs of all propositions and lemmas.

2 Related literature

Our paper is closest to papers that explore motives to tax capital income if the Atkinson and Stiglitz (1976) theorem breaks down. In particular, if preferences are identical but non-separable, taxes (or subsidies) on capital can be optimal income to alleviate distortions of labor taxes on labor supply (Corlett and Hague, 1953; Atkinson and Stiglitz, 1976; Erosa and Gervais, 2002; Conesa, Kitao, and Krueger, 2009; Jacobs and Boadway, 2014). Moreover, individuals may differ in their preference to save. Banks and Diamond (2010) cite ample empirical evidence suggesting that preferences to save (i.e., discount rates) are heterogeneous and correlated with earning ability in labor income. Consequently, capital incomes differ between individuals – conditional on their labor income – and positive taxes on capital income are optimal (Mirrlees, 1976; Saez, 2002; Diamond and Spinnewijn, 2011). Second, individuals can also differ in their endowments and inheritances. Piketty (2014) presents evidence that inheritances are indeed positively associated with labor earnings. If endowments or inheritances correlate positively with labor earnings but cannot be taxed, then taxing capital income is socially optimal, even if weakly separable preferences are assumed (Cremer, Pestieau, and Rochet, 2001). Third, if inheritances are an additional source of wealth besides their labor income, and if the government is constrained to using linear taxes on labor income and inheritances, then it is optimal to tax inheritances (Piketty and Saez, 2013). Fourth, individuals might derive direct utility from wealth, so it can be optimal to tax capital income at positive rates for purely redistributive reasons (Saez and Stantcheva, 2016).7

Recently, a number of papers also analyze optimal taxation in models with heterogeneous capital returns. In particular, Gahvari and Micheletto (2016) analyze optimal capital taxes in the two-type optimal tax framework of Stiglitz (1982), where capital incomes are determined by labor earnings to model the notion that high-earning individuals have better access to financial markets. Optimal taxes on capital income are shown to be positive if capital income correlates positively with labor income. Kristjánsson (2016) also analyzes a two-type framework where heterogeneity in capital returns originates from differences in entrepreneurial investment effort and ability. Optimal taxes on either

---

7 Like this paper, Saez and Stantcheva (2016) argue that taxing capital income could be desirable if returns are heterogeneous. However, in their framework taxes on capital income are positive as well if returns are homogeneous. Hence, it is not a priori clear what return heterogeneity does to optimal taxes on capital income – compared to the case with homogeneous returns. In contrast, our analysis is based on the conventional life-cycle model. Taxes on capital income are optimally zero in the absence of return heterogeneity due to uniform and weakly separable preferences.
capital income or wealth are shown to be positive as well. Finally, Guvenen, Kambourov, and Kuruscu (2017) explore applied DSGE-OLG-models in which individuals have heterogeneous capital returns due to entrepreneurial efforts. They find that optimal linear wealth taxes yield higher social welfare than optimal linear taxes on capital income, as taxes on capital income distort entrepreneurial efforts more heavily.

We contribute to this recent literature by analyzing returns to capital that are a function of both ability and savings, rather than a function of labor income or entrepreneurial effort. This allows us to study the role of ability rents and scale effects in capital returns, the latter of which have not been analyzed before. Moreover, we analyze models with a continuum of types as in Mirrlees (1971), rather than two types to study non-linear taxes on capital income and to gain more insight into the shape of optimal non-linear schedules on capital income. This also allows us to relate the optimal tax rates to key statistics like the elasticities of the tax bases and the income distribution. Moreover, we are able to derive conditions for a Pareto efficient dual income tax with separate rates on labor and capital income.

Our paper is also related to a large literature exploring the conditions under which optimal taxes on capital income could be positive for efficiency reasons. Chamley (1986), Judd (1985), Straub and Werning (2014), and Jacobs and Rusu (2017), explore dynamic representative-agent or two-class infinite-horizon models and find that taxes on capital income may be zero in the steady state. Ordover and Phelps (1979), Atkinson and Sandmo (1980) and King (1980) analyze OLG-models and show that the optimal tax on capital income may be non-zero to correct dynamic inefficiencies in capital accumulation. Jacobs and Bovenberg (2010) demonstrate that taxes on capital income are optimal to alleviate labor-tax distortions on human capital formation. Aiyagari (1994), Hubbard and Judd (1986), Diamond and Mirrlees (1978), Golosov, Kocherlakota, and Tsyvinski (2003) and Jacobs and Schindler (2012) show that optimal taxes on capital income are positive if capital and insurance markets are missing. Christiansen and Tuomala (2008) and Reis (2010) derive that positive taxes on capital income are optimal to avoid tax arbitrage between labor and capital income. We refer the reader to the overviews provided in Banks and Diamond (2010), Diamond and Saez (2011), and Jacobs (2013) for extensive discussions of these arguments.

Finally, although our model abstracts from risk in capital returns, our paper is related to a literature that finds that capital income should (optimally) be taxed in models with risk and portfolio choice. Varian (1980) shows that optimal taxes on capital income are positive if returns to saving feature idiosyncratic risk and tax revenues can be returned in lump-sum fashion. Gordon (1985) and Spiritus and Boadway (2017) study optimal taxation of risk-free and risky assets. They find that taxes on capital income yield no insurance gains, and can thus be arbitrarily set, if there is only aggregate risk in capital returns and all revenues have to be returned in state-contingent lump-sum transfers. Schindler (2008) studies the optimal tax on the risk premium if there is aggregate risk in capital returns and tax revenue is used to finance state-contingent public goods. He derives that the optimal tax balances the risk of private consumption against the risk of public good provision. If capital returns feature both idiosyncratic and systematic risk components, Spiritus and Boadway (2017) find that the optimal tax on the risk premium is generally positive with both state-contingent lump-sum rebates and state-contingent public good provision. Hence, if we would allow risk or portfolio choice, our case for positive taxes on capital income would only be strengthened.
3 Model

3.1 Individual behavior

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 \( n \). Individuals are assumed to live for two periods. In the first period, individual \( n \) supplies labor \( l^n \) and earns labor income \( z^n \equiv nl^n \). He spends his first-period income on taxes on labor income \( T^n \), consumption \( c_1^n \), and savings \( a^n \). Thus, we can write first-period consumption as:

\[
(1) \quad c_1^n = z^n - T^n - a^n.
\]

Savings yield 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 a number of plausible microfoundations of return heterogeneity related to closely-held assets and scale economies in wealth investment. Taxes on capital income are denoted by \( \tau^n \), and second-period consumption equals the sum of savings and net-of-tax capital income:

\[
(2) \quad c_2^n = a^n + y(a^n, n) - \tau^n.
\]

Capital income is assumed to be deterministic. As mentioned in the Introduction, heterogeneity in returns may partly originate from varying exposure to idiosyncratic risk. We nevertheless abstract from risky returns for two main reasons. First of all, empirical evidence shows that people significantly differ even in the rates of return on their relatively risk-free asset (Fagereng et al., 2016). Second, it is already well known that the existence of idiosyncratically risky returns calls for positive taxes on capital income as a form of social insurance (e.g., Varian, 1980; Spiritus and Boadway, 2017).

\( T^n \) is a nonlinear tax function of labor income \( z^n \), and \( \tau^n \) is a nonlinear 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 helps us define behavioral elasticities and social welfare weights.\(^8\) We write the tax schedules as the following functions:

\[
(3) \quad T^n = T(z^n, \rho^T, \sigma^T) = \tilde{T}(z^n) + \rho^T + \sigma^T z^n;
\]

\[
(4) \quad \tau^n = \tau(y^n, \rho^\tau, \sigma^\tau) = \tilde{\tau}(y^n) + \rho^\tau + \sigma^\tau y^n;
\]

where \( \rho^T \) and \( \rho^\tau \) are parameters that shift the intercepts of the tax schedules, and \( \sigma^T \) and \( \sigma^\tau \) 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 nonlinear 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:

\[
(5) \quad U^n = u(c_1^n, c_2^n) - v(z^n/n).
\]

\(^8\)It is not uncommon to parameterize nonlinear tax schedules for the derivation of comparative statics, see, e.g., Christiansen (1981); Immervoll et al. (2007); Jacquet, Lehmann, and Van der Linden (2013); Gerritsen (2016).
Utility of consumption $u(\cdot)$ is increasing, concave, and twice differentiable. Disutility of work $v(\cdot)$ is increasing, strictly convex and twice 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 yield the following first-order conditions:

$$
\frac{u'(z^n/n)}{u_1(c^n_1, c^n_2)} = (1 - T'(z^n, \rho^T, \sigma^T))n,
$$

$$
\frac{u_2(c^n_1, c^n_2)}{u_1(c^n_1, c^n_2)} = \frac{1}{1 + (1 - \tau(y(a^n, n), \rho^\tau, \sigma^\tau))y_a(a^n, n)} = \frac{1}{R^n}.
$$

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 $y_a(\cdot)$ is the marginal return to savings. Eq. (6) shows that the marginal rate of substitution between (first-period) consumption and leisure must equal the marginal net-of-tax wage rate. Eq. (7) shows that the marginal rate of substitution of first-period consumption for second-period consumption must equal the individual’s discount factor. We define the inverse of the discount factor – or one plus the net-of-tax rate of return – as $R^n \equiv 1 + (1 - \tau')y_a$.

We impose a number of assumptions that help us derive the optimal non-linear tax schedules. We require that both tax schedules are continuous and twice differentiable. To ensure smooth behavioral responses to marginal tax changes, i.e., to rule out bunching after a tax change, we assume that second-order conditions are always satisfied so that eqs. (6) and (7) describe utility-maximizing choices. Hence, the first-order conditions describe a unique, global maximum for utility. If the second-order condition for utility maximization are satisfied, the equilibrium values of both tax bases $y^n$ and $z^n$ are monotonically increasing in ability $n$. See also Assumption 2 in Jacquet and Lehman (2017).

### 3.2 Behavioral elasticities

Behavioral elasticities of the tax bases play an important role in optimal tax formulas. 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 $a^n = a(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$. As capital income is a function of savings and ability, we can write equilibrium capital income as a function of the same arguments $y^n = y(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$. The combination of both first-order conditions in eqs. (6) and (7), together with the definitions of first- and second-period consumption, determines labor income as a function of tax parameters and ability. This allows us to write equilibrium labor income as $z^n = z(\rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$.

We define the compensated elasticity of labor income with respect to the net-of-tax rate for each

---

9In what follows, we suppress function arguments for brevity unless this is likely to cause confusion.

10We denote equilibrium functions for the tax bases with a tilde. We do this to distinguish equilibrium capital income $\tilde{y}(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$ from capital income as a function of savings and ability $y(a^n, n)$. 


individual $n$ as:

$$
e^n_z \equiv -\left( \frac{\partial \tilde{z}(\rho^T, \rho^r, \sigma^T, \sigma^r, n)}{\partial \sigma^T} - \frac{n}{\partial \rho^T} \right) \frac{1 - T'}{z^n}.
$$

The term within brackets gives the Slutsky decomposition of the compensated response in labor income to an increase in marginal taxes. The first term is the total effect of an increase in the tax on labor income that is proportional in income. The second term is the effect of a compensating reduction in the intercept of the tax schedule. Multiplied by $-(1 - T')/z^n$, it 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. An exogenous change in the net-of-tax rate leads to a change in labor income, which leads to an endogenous change in the net-of-tax rate, which further affects labor income. The elasticity in eq. (8) is defined to take into account the total effect on labor income.\textsuperscript{11}

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

$$
e^n_y|z \equiv -\left( \frac{\partial \tilde{y}(z^n, \rho^T, \rho^r, \sigma^T, \sigma^r, n)}{\partial \sigma^r} - \frac{n}{\partial \rho^r} \right) \frac{R^n}{y^n}.
$$

Again, the term within brackets is the Slutsky decomposition of the compensated response of capital income to an increase in marginal taxes. A unit change in the marginal tax on capital income $y_a$ implies a change in the rate of return $R^n = 1 + (1 - \tau')y_a$. Thus, multiplying the compensated response in capital income by $R^n/y_a$, and dividing by $y^n$, yields the percentage change in the capital income of individual $n$ if the rate of return $R^n$ is exogenously raised by one percent. The elasticity takes into account the total effect on capital income. Thus, it incorporates the fact that an exogenous change in the interest rate $R^n$ may lead to further endogenous changes in $R^n$, both of which affect capital income. 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:

$$
\xi^n_z \equiv \frac{\partial \tilde{z}(\rho^T, \rho^r, \sigma^T, \sigma^r, n)}{\partial n} \frac{n}{z^n}.
$$

$$
\xi^n_y|z \equiv \frac{\partial \tilde{y}(z^n, \rho^T, \rho^r, \sigma^T, \sigma^r, n)}{\partial n} \frac{n}{y^n}.
$$

The first elasticity $\xi^n_z$ 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.\textsuperscript{11}

In terms of Jacquet and Lehmann (2017), $e^n_z$ is a ‘total elasticity’ rather than a ‘direct elasticity.’ Technically, this is because the elasticity measures the effect on labor income of a given change in the tax parameters $\sigma^r$ and $\rho^r$ rather than a given change in the marginal tax rate $T'(z, \rho^r, \sigma^T)$. 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).
4 Two microfoundations of return heterogeneity

4.1 Closely-held assets

It is instructive to consider two plausible microfoundations for capital income $y(a^n, n)$ that could generate heterogeneity in rates of return. First, consider an economy in which all individuals have equal access to some commonly traded asset that is traded at international capital markets at some exogenous rate of return $r$. On top of that, individuals may invest in a closely held asset. This could be interpreted as entrepreneurial investment. Individual $n$ invests $b^n$ in the closely-held asset and thus $a^n - b^n$ in the commonly traded asset. The gross return to entrepreneurial investment is a function of invested capital and ability, and given by $\pi^n = \pi(b^n, n)$. We assume that it is subject to decreasing returns to capital ($\pi_b > 0$ and $\pi_{bb} < 0$) and increasing in ability ($\pi_n > 0$). The latter assumption reflects the idea that high ability helps to find and select successful investment opportunities.

Capital income is now given by:

$$y_n = r(a^n - b^n) + \pi(b^n, n) = ra^n + (\pi(b^n, n) - rb^n).$$

Individuals spread 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. (12) 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 back into the equation for capital income (12) yields:

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

Hence, the general formulation $y^n = y(a^n, n)$ captures the special case of entrepreneurial investments. Notice that, under this microfoundation, capital income is linear in savings and increasing in ability: $y_a = r$ and $y_n > 0$.

4.2 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 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.

To capture this in our model, assume that individuals could spend some amount $x^n$ of their savings on some fixed investment costs – search costs, fees, costs of obtaining financial know-how – that improves their rates of return. This leaves an amount $a^n - x^n$ to be invested at a rate of return $r(x^n) \geq 0$ with

\footnote{Both follow from the partial derivatives of $y(a, n)$ in eq. (13). $y_a = r$ follows trivially. Application of the envelope theorem yields $y_n = \pi_n \geq 0$.}
\( r'(x^n) \geq 0 \). We assume that all investment costs are tax deductible.\(^{13}\) Taxable capital income is then given by:

\[
y^n = r(x^n)(a^n - x^n) - x^n.
\]

(14)

Individuals spend their savings on investment costs as long as it leads to an increase in total capital income. Maximizing \( y^n \) in eq. (14) with respect to \( x^n \) yields \( r'(x^n)(a^n - x^n) = 1 + r(x^n) \).\(^{14}\) 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 back into the expression for capital income yields:

\[
y^n = y(a^n, n) = r(x(a^n))(a^n - x(a^n)) - x^n(a^n).
\]

(15)

Hence, the general formulation \( y^n = y(a^n, n) \) could also capture 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 \( y_n = 0 \).\(^{15}\)

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 usefully be interpreted as the costs of entering a specific financial market in which assets pay off a return \( r(x) \). Thus, individuals with different levels of wealth effectively invest in different financial markets. As a result, there is not one single financial market that equates marginal rates of transformation. This means that there are potential Pareto-improving transactions that are not executed.

To see this, imagine that a poor, low-return individual could lend funds to a wealthy, high-return individual at some intermediate interest rate. Such transaction would be mutually beneficial because the wealthy individual could invest the resources in a higher-yielding asset, while the poor individual could divest from his lower-yielding asset. Thus, implicit in the micro-foundation is a market failure that keeps the poor from accessing the higher-yielding investment opportunities of the rich.

5 Optimal taxation

5.1 Social welfare and government budget constraints

The social planner is assumed to be utilitarian. Thus, social welfare is written as the sum of all individual utilities:

\[
W = \int_0^\infty U^n f(n)dn.
\]

\(^{13}\)Typically, investment funds subtract their fees from the payout to the participants. This effectively makes the investment fees tax deductible for the owner of the wealth.

\(^{14}\)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 some 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, the second-order condition is always satisfied.

\(^{15}\)This follows from the partial derivatives of eq. (15). \( y_n = 0 \) follows trivially. Application of the envelope theorem yields \( y_a = r(x(a^n)) \geq 0 \) and, hence, \( y_{aa} = r'x' \geq 0 \).
It is straightforward to introduce stronger redistributional social preferences, for example with individual-specific Pareto weights or by taking a concave transformation of utility.

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:

\[
B_1 = \int_0^{\infty} T(z^n, \rho^T, \sigma^T) f(n) dn - g_1,
\]

\[
B_2 = \int_0^{\infty} \tau(y^n, \rho^\tau, \sigma^\tau) f(n) dn - g_2,
\]

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

Instead of assuming exogenously fixed government assets, we could alternatively assume that the government has access to the same investment technology as individuals. In the first microfoundation with commonly traded assets, this would imply that the government could borrow and lend at the same constant marginal rate of return as every individual. As a result, government debt would be completely neutral so that the optimal net asset position of the government becomes indeterminate. Thus, assuming that the government’s net asset position is fixed at some exogenously given level is entirely innocuous in case of the first microfoundation.

In the second microfoundation with scale economies, the government could presumably invest at a marginal rate of return that is greater than that of every individual given its size. In that case, the government would want to invest on behalf of individuals by raising lump-sum taxes and government assets in the first period and redistribute returns by lowering lump-sum taxes in the second period. In reality, we believe that the government is limited in the extent to which it can invest resources on behalf of its citizens at superior rates of return. First of all, there may be inefficiencies associated with a large public investment portfolio, generating declining rates of return on government assets. Second, raising lump-sum taxes in the first period to invest on behalf of individuals only works if the poor could borrow against future government transfers. In reality, poor households may face significant borrowing constraints. Third, there may be political-economy reasons why politicians should not be entrusted with large public investment funds. Instead of microfounding these restrictions on government assets, we decide to simply assume that the net asset position of the government is exogenously fixed.

We denote the shadow prices of first- and second-period government revenue by \(\lambda_1\) and \(\lambda_2\), so that the social planner’s objective function can be written as:

\[
\mathcal{L} = \frac{1}{\lambda_1} W + B_1 + \frac{1}{\lambda_1/\lambda_2} B_2.
\]

Thus, the government discounts future tax revenue at a rate \(\lambda_1/\lambda_2\).

### 5.2 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:

\[
E^n_T \equiv -T' \left( \frac{dz^n}{d\sigma^T} - z^n \frac{dz^n}{d\rho^T} \right) - \frac{\tau'}{\lambda_1/\lambda_2} \left( \frac{dy^n}{d\tau^T} - z^n \frac{dy^n}{d\rho^T} \right),
\]

\[
E^n_\tau \equiv -T' \left( \frac{dz^n}{d\sigma^\tau} - z^n \frac{dz^n}{d\rho^\tau} \right) - \frac{\tau'}{\lambda_1/\lambda_2} \left( \frac{dy^n}{d\tau^\tau} - y^n \frac{dy^n}{d\rho^\tau} \right).
\]

An increase in marginal taxes potentially affects both tax bases, thereby affecting both first- and second-period revenue. Eq. (20) 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. (21) 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 \(\alpha^n_1\) and \(\alpha^n_2\):

\[
\alpha^n_1 \equiv \frac{u_1}{\lambda_1} - T' \frac{dz^n}{d\rho^T} - \frac{\tau'}{\lambda_1/\lambda_2} \frac{dy^n}{d\rho^T},
\]

\[
\alpha^n_2 \equiv \frac{u_2}{\lambda_1} - T' \frac{dz^n}{d\rho^\tau} - \frac{\tau'}{\lambda_1/\lambda_2} \frac{dy^n}{d\rho^\tau}.
\]

The social welfare weights consist of the utility 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.

5.3 Optimal tax schedules

A benevolent government maximizes the Lagrangian in eq. (19) by optimally setting the nonlinear taxes on labor and capital income. We approach this problem by using the tax-perturbation approach, which was pioneered by Saez (2001) and has more recently been extended and amended by Golosov, Tsyvinski, and Werquin (2014), Gerritsen (2016), and Lehmann et al. (2018). In particular, we consider small changes in the marginal tax rates on labor and capital income over small intervals of the tax bases. In the optimum, such tax perturbations should have no effect on social welfare. Denote the density of labor income by \(h(z^n)\) and the density of capital income by \(g(y^n)\). 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 \(y^n\):

\[
E^n_T h(z^n) = \int_{z^n}^{\infty} (1 - \alpha^n_1) h(z^m) dz^m,
\]

\[
E^n_\tau g(y^n) = \int_{y^n}^{\infty} \left( \frac{1}{\lambda_1/\lambda_2} - \alpha^n_2 \right) g(y^m) dy^m.
\]

**Proof.** See Appendix.
The conditions in Lemma 1 are intuitively straightforward: the marginal deadweight costs of raising marginal taxes should equal their marginal redistributional gains. A small increase of the marginal tax on labor income rate around $z^n$ 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. (24). 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. (24). In the same spirit, the left-hand side of eq. (25) gives the marginal excess burden of raising the marginal tax rate around $y^n$ and the right-hand side gives its redistributional gains.

### 5.4 Optimal taxation of labor income

Lemma 1 implicitly expresses optimal tax schedules in terms of marginal excess burdens and redistributional gains. To gain more insight into the shape of the optimal tax schedules, we explicitly 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:

\[
\frac{T'(z)}{1-T'(z)} + \frac{y_a \tau'(y)}{\lambda_1/\lambda_2} = \frac{1}{e_z} \cdot \frac{1-H(z)}{zh(z)} \cdot (1-\bar{\alpha}_{1|z}),
\]

where $s \equiv \frac{1}{1-H(z)} \frac{\partial z}{\partial \sigma}$ is the propensity to save out of net income, $H(z)$ is the cumulative distribution function of labor income, and $\bar{\alpha}_{1|z} \equiv \int_0^\infty (1-\alpha^n)h(z^n)dz^n/(1-H(z))$ is the average first-period social welfare weight of individuals that earn more than $z$.

**Proof.** Recall that $z^n = \tilde{z}(\rho^T, \rho^*, \sigma^T, \sigma^*, n)$. Thus, we can write $\frac{dz^n}{d\sigma^T} = \frac{\partial \tilde{z}}{\partial \sigma^T}$ and $\frac{dz^n}{d\rho^T} = \frac{\partial \tilde{z}}{\partial \rho^T}$. Second, recall that $y^n = y(a^n, n)$ and $a^n = \tilde{a}(z^n, \rho^T, \rho^*, \sigma^T, \sigma^*, n)$. Moreover, a compensated change in the marginal tax rate on labor income only affects savings through labor income, i.e.: $\frac{\partial z^n}{\partial \sigma^T} - z^n \frac{\partial \tilde{a}}{\partial \rho^T} = 0$. We can therefore write $\frac{dy^n}{d\sigma^T} - z^n \frac{\partial y^n}{\partial \sigma^T} = y_a \frac{\partial \tilde{a}}{\partial \sigma^T} \left( \frac{\partial z^n}{\partial \rho^T} - z^n \frac{\partial \tilde{z}}{\partial \rho^T} \right) = (1-T')y_a s^n \left( \frac{\partial z^n}{\partial \rho^T} - z^n \frac{\partial \tilde{z}}{\partial \rho^T} \right)$. Substitute these expressions and the elasticity of eq. (8) into the definition of the excess burden in eq. (20), and substitute this into the optimality condition in eq. (24). Rearranging yields eq. (26). \qed

The left-hand side of eq. (26) gives the tax wedge on net labor income for an individual with income $z$. To see this, consider a unit increase in net-of-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 is discounted at a rate $\lambda_1/\lambda_2$. The right-hand side of eq. (26) is the standard expression for the optimal tax wedge on labor, see also Mirrlees (1971); Diamond (1998); Saez (2001). The optimal tax wedge on labor income is decreasing in the elasticity of labor income at $z$, the relative hazard rate of the income distribution $zh(z)/(1-H(z))$, and the average of the social-welfare weights of individuals who earn more than $z$. The only material difference with, e.g., Saez (2001), is that the tax wedge on labor income consists not only of 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 tax wedge on labor income merely consist of the marginal tax rate on labor income.

5.5 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 in which return heterogeneity originates from closely-held assets. We then discuss the case 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.

5.5.1 Closely-held assets \((y_{aa} = 0, y_n \geq 0)\)

Recall from Section 4 that one possible microfoundation for capital income \(y(a, n)\) reflects the existence of privately-held assets with decreasing returns that are positively related to ability. Together with a commonly traded asset, this microfoundation assures that capital income is linear in savings and increasing in ability: \(y_{aa} = 0\) and \(y_n \geq 0\). Under these assumptions, we obtain the optimal tax on capital income as described by the following Proposition.

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

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

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

**Proof.** See Appendix.

Eq. (27) implicitly equates the deadweight loss of a small increase in the tax on capital income with the deadweight loss of a small increase in the tax on labor income, both for a given amount of redistribution. On the margin, both taxes on labor and capital income can generate the same redistribution from rich to poor. In the optimum, the government must be indifferent between using either tax instrument. Hence, for the same redistribution of income, the government optimally equates the marginal deadweight losses of the two taxes on labor and capital income. Both taxes distort labor supply, but taxes on capital income additionally distort savings. The left-hand side of eq. (27) gives the savings distortion of a tax on capital income. It equals the tax wedge on savings multiplied by the conditional elasticity of capital income with respect to the tax rate on capital income. The right-hand side gives the distortions on labor supply of a tax on labor income relative to a tax on capital income. The term in brackets is the labor-supply distortion of a tax on labor income, which equals the tax wedge on labor income multiplied by the elasticity of labor income. The labor distortions are multiplied the ratio of ability elasticities \(\xi_{y|z}/\xi_z\), which indicates the degree to which the tax on capital income distorts labor supply less than the tax on labor income. Eq. (27) shows that the optimal tax on capital income is positive if and only if this ratio is positive, which is the case if, conditional on labor income, capital income is increasing in ability \((y_n(a, n) > 0)\).

Intuitively, the ratio \(\xi_{y|z}/\xi_z\) captures the extent to which ability correlates more strongly with capital income than with labor income. The second-best nature of optimal taxation originates from
the fact that the government cannot observe and therefore cannot tax ability. Instead, the government wants to tax those tax bases that provide informational content on ability. One could view each tax base as a function of hidden ability and hidden labor efforts. The stronger a tax base correlates with ability, the more the tax resembles a non-distortionary tax on ability. On one extreme, if a tax base only correlates with ability but not with effort, then the tax redistributes without distortions and could be used to achieve first best. On the other extreme, if a tax base does not correlate with ability at all, then the tax produces distortions without any distributional benefits. Hence, a positive ratio $\xi_{y|x}/\xi_{z}$ implies that a tax on capital income can redistribute more income for less labor-supply distortions than a tax on labor income. As a result, a strictly positive tax on capital income is optimal even if it distorts savings decisions.

The optimality condition in eq. (27) 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 $\xi_{y|x}/\xi_{z}$. Second, the optimal tax rate on capital income is decreasing in the conditional elasticity of capital income with respect to the interest rate $e_{y|x}$. 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 optimal tax on capital income tracks the optimal tax wedge on labor income. As the optimal tax wedge on labor income tends to be U-shaped in income, we expect the optimal tax on capital income to be U-shaped in income as well (e.g., Saez (2001)). In Section 6, we confirm this by numerically simulating the optimal tax schedules for empirically realistic return heterogeneity.

Proposition 2 is closely related to a number of earlier contributions to the literature on optimal commodity taxation and the literature on optimal taxation of capital income. Atkinson and Stiglitz (1976) show that the government should not use commodity taxes 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) noted 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 demand change for given income and labour 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). 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 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 more closely related to Cremer, Pestieau, and

\[ \frac{\xi_{y|x}}{\xi_{z}} = \frac{\partial y/\partial n}{\partial z/\partial n} z = \left( \frac{dy/\partial n}{\partial z/\partial n} - \frac{\partial y}{\partial z} \right) \frac{z}{y} \]

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.
Rochet (2001), who find that taxes on capital income are desirable if endowments are increasing with ability.

Furthermore, the results in Proposition 2 are related to two independent studies by Gahvari and Micheletto (2016) and Kristjánsson (2016). They also study return heterogeneity and show that positive taxes on capital income are optimal if rates of return are increasing in ability for given labor income. Our result differs from theirs in a number of ways. They focus on an economy with two types of individuals, whereas we study a continuum of types. This allows us to derive an optimality condition that is written in terms of a limited number of sufficient statistics. It also allows us to be explicit on how optimal marginal taxes on capital income vary over the income distribution. We furthermore simulate optimal taxes on capital income for a realistic income distribution in Section 6 below.

5.5.2 Scale economies in wealth investment (\(y_{aa} \geq 0, y_n = 0\))

The second microfoundation for capital income \(y(a, n)\) that we considered in Section 4 reflects the existence of economies of scale in wealth management: \(y_{aa} \geq 0\) and \(y_n = 0\). With \(y_n = 0\), ability is equally strongly correlated to labor and capital income. As a result, for the same amount of redistribution, marginal taxes on labor and capital income lead to the same labor-supply distortions. Nevertheless, the following Proposition establishes the optimality of positive taxes on capital income as long as rates of return are increasing in savings.

**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), the optimal marginal tax rate on capital income for every level of capital income \(y\) is implied by:

\[
\tau'(y)y_a = \frac{1 + \bar{y}_a}{1 - G(y)} \cdot \frac{1 - G(y)}{yg(y)} \cdot \frac{\bar{y}_{a|y} - \bar{y}_a}{1 + \bar{y}_a} \geq 0,
\]

where \(G(y)\) is the cumulative distribution function of capital income, \(\bar{y}_{a|y} \equiv \int_y^\infty y_a(a^n, n)g(y^n)dy^n/(1 - G(y))\) is the average marginal rate of return for individuals whose capital income is more than \(y\), and \(\bar{y}_a \equiv \bar{y}_{a|0} = \int_0^\infty y_a(a^n, n)g(y^n)dy^n\) is the average marginal rate of return for all individuals.

**Proof.** See Appendix.

Heterogeneity in marginal rates of return implies an absence of a financial market that equates marginal rates of intertemporal transformation. Poor low-return individuals are unable to invest in the relatively high-yielding assets of wealthy high-return individuals. Even though the government also cannot borrow or invest on behalf of its citizens – its debt is assumed to be fixed – it can use its tax policy to partially correct for the market failure. Indeed, by shifting taxes away from labor income towards capital income, the government can replicate the missing financial transactions between the wealthy and the poor. A budget-neutral reduction in marginal taxes on labor income transfers first-period resources from the poor to the rich. A budget-neutral increase in marginal taxes on capital income transfers second-period resources back from the rich to the poor. The combination of these two budget-neutral tax reforms effectively forces the wealthy to invest on behalf of the poor.

An alternative but equivalent interpretation of this result is as follows. 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

\[
\tau'(y)y_a = \frac{1 + \bar{y}_a}{1 - G(y)} \cdot \frac{1 - G(y)}{yg(y)} \cdot \frac{\bar{y}_{a|y} - \bar{y}_a}{1 + \bar{y}_a} \geq 0,
\]

where \(G(y)\) is the cumulative distribution function of capital income, \(\bar{y}_{a|y} \equiv \int_y^\infty y_a(a^n, n)g(y^n)dy^n/(1 - G(y))\) is the average marginal rate of return for individuals whose capital income is more than \(y\), and \(\bar{y}_a \equiv \bar{y}_{a|0} = \int_0^\infty y_a(a^n, n)g(y^n)dy^n\) is the average marginal rate of return for all individuals.

**Proof.** See Appendix.

Heterogeneity in marginal rates of return implies an absence of a financial market that equates marginal rates of intertemporal transformation. Poor low-return individuals are unable to invest in the relatively high-yielding assets of wealthy high-return individuals. Even though the government also cannot borrow or invest on behalf of its citizens – its debt is assumed to be fixed – it can use its tax policy to partially correct for the market failure. Indeed, by shifting taxes away from labor income towards capital income, the government can replicate the missing financial transactions between the wealthy and the poor. A budget-neutral reduction in marginal taxes on labor income transfers first-period resources from the poor to the rich. A budget-neutral increase in marginal taxes on capital income transfers second-period resources back from the rich to the poor. The combination of these two budget-neutral tax reforms effectively forces the wealthy to invest on behalf of the poor.

An alternative but equivalent interpretation of this result is as follows. 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 individual behavior for now and consider two individuals: a poor individual with a zero marginal rate of return \((y_a = 0)\) and a rich individual with a strictly 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, for the same utility gains of the poor, the government could reduce the utility losses of the rich by redistributing in the second period instead of the first period. That is, redistribution of second-period resources Pareto dominates redistribution of first-period resources. Because taxes on capital income are typically levied later in life than taxes on labor income, this yields an argument in favor of positive taxes on capital 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 \((y_{aa} > 0 \text{ and } y_n = 0)\) 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 particular assumption that taxes on labor and capital income are levied at the same point in time. Proposition 3 shows that their claim is no longer valid when the government realistically levies taxes on capital income later in life. In that case, the government should tax capital income when rates of return are increasing in savings.

The left-hand side of eq. (28) is the tax wedge on savings for individuals with capital income \(y\). In the optimum, this wedge equals the product of three terms: the inverse of the conditional elasticity of capital income \(1/e_{y|z}\); the inverse hazard rate of the distribution of capital income \((1−G(y))/yg(y)\); and the discounted difference between the average marginal rate of return of the rich and the average marginal rate of return of the total population \((\bar{y}_a|y − \bar{y}_a)/(1 + \bar{y}_a)\). First, a larger conditional elasticity implies that a tax on capital income distorts saving decisions more heavily. As a result, the optimal tax on capital income is decreasing in this elasticity. Second, a large inverse hazard rate at \(y\) implies that the marginal tax on capital income raises revenue from a large proportion of people \(1−G(y)\) relative to the concentration of capital income that it distorts \(yg(y)\). As a result, the optimal tax on capital income is increasing in the inverse hazard rate. Third, an increase in the marginal tax on capital income around \(y\) raises the second-period tax burden of individuals with capital income above \(y\). This causes them to increase their savings for purposes of consumption smoothing, earning an average return of \(\bar{y}_a|y\). At the same time, a budget-neutral reduction in the intercept of the tax schedule reduces savings of the entire population, on which they would otherwise obtain an average return of \(\bar{y}_a\). Thus, a budget-neutral increase in the marginal tax on capital income around \(y\) yields a discounted increase in total returns of \((\bar{y}_a|y − \bar{y}_a)/(1 − \bar{y}_a)\). The larger this gain, the more beneficial the tax reform, and thus the higher the optimal marginal tax rate on capital income.

One attractive feature of the optimality condition of eq. (28) is that it expresses the optimal tax

---

17 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 propagate 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, the government effectively wants to stimulate future labor supply by reducing the tax on future labor income. However, if it is unable to differentiate taxes on current and future labor income, a tax on capital income may stimulate future labor supply instead. This argument does not play a role in our model because individuals only supply labor in one period.

18 This is the same reason for which the optimal marginal tax rate on labor income is increasing in the inverse hazard rate of the distribution of labor income, see Proposition 1 and, e.g., Saez (2001).
on capital income exclusively in terms of sufficient statistics with clear empirical counterparts. The conditional elasticity of capital income could be estimated by using plausibly exogenous variation in taxes on capital income – even though currently empirical evidence on this particular elasticity is still relatively scarce. 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., 2016). The only empirical matter on which we currently 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 heterogeneity in closely-held assets – as studied in Proposition 2.

Naturally, even if we can empirically identify all the sufficient statistics in eq. (28), these statistics are typically endogenous to the tax system itself. Thus, whereas the statistics in eq. (28) refer to their values in the tax optimum, we can only observe their values at the current tax system. This makes it hard to directly use eq. (28) to quantify the optimal tax on capital income. Fortunately, this is less of a problem for the optimal tax rate at the top of the income distribution, as is shown in the following Corollary.

**Corollary 1.** Assume that capital income is convex in wealth but not directly affected by ability ($y_{aa} ≥ 0$ and $y_n = 0$ for all individuals). If the tail of the distribution of capital income follows a Pareto distribution, and if the elasticity of capital income and the marginal rate of return on savings converge to the constants $\hat{\epsilon}|_{z}$ and $\hat{y}_a$ for high income levels, then the optimal tax rate on capital income at the top of the income distribution is constant and given by:

$$\tau'(\hat{y}) = \frac{1}{\epsilon|_{z}} \cdot \frac{1}{p} \cdot \left(1 - \frac{\hat{y}_a}{\hat{y}_a}\right),$$

where a ‘hat’ denotes variables that refer to individuals in the top of the distribution of capital income, $G(y)$ is the cumulative distribution function of capital income, and $p = \frac{1-G(y)}{y_9(y)}$ is the Pareto parameter of the top of the distribution of capital income.

**Proof.** Substituting $y_a = \tilde{y}_{a|y} = \tilde{y}_a$, $\epsilon|_{z} = \hat{\epsilon}|_{z}$, and $\frac{1-G(y)}{y_9(y)} = p$ into eq. (28) yields eq. (29).

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 and the thinness of the income distribution’s tail. Furthermore, it is increasing in the marginal rate of return on savings at the top relative to the average marginal rate of return.

Eq. (29) allows us to make some back-of-the-envelope calculations of optimal top tax rates on capital income. Fagereng et al. (2016) find that the average Norwegian in 2013 obtained an average rate of return of about 3.7%, whereas the average rate of return of the wealthiest decile was about 2 percentage points higher. This implies a value for $\tilde{y}_a/\hat{y}_a$ of 3.7/5.7 = 0.65. If the tail of the distribution of capital income is comparable to that of broader income measures, $p = 2$ would be a conservative estimate of the Pareto parameter of the Norwegian capital-income distribution (e.g., Aaberge and Atkinson, 2010). Finally, there are few good estimates on the elasticity of capital income, but a conservative estimate may be $\hat{\epsilon}|_{z} = 0.4$ – which is in line with elasticities of taxable income (Saez, Giertz, and Slemrod, 2012). Taken together, these values imply an optimal top tax rate on capital income of 43%. This tax

---

19 This holds especially when compared to the abundance of estimates of the elasticity of labor income. See Zoutman (2015) for a recent study on the return elasticity of wealth.
rate is substantial, and not much less than the top tax rate on labor income in many countries (e.g. OECD, 2018).

5.5.3 The general case

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

\[
R^n = R(y^n, n) = 1 + (1 - \tau'(y^n))y_a(a(y^n, n), n).
\]

Using this notation, we can now present an expression for the optimal tax on capital income when 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:

\[
\left( \frac{\tau'(y) y_a}{\lambda_1/\lambda_2} \right) e_{y|z} = \left( \frac{\xi_y}{\xi_z} \right) \cdot \left( \frac{T'(z)}{1 - T'(z)} + \frac{s y_a \tau'(y)}{\lambda_1/\lambda_2} \right) e_z + \frac{1 - G(y)}{y g(y)} \cdot \left( \frac{\bar{y}_{a|y} - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} \right) - \frac{1}{y g(y)} \cdot \int_n^\infty \left( \frac{\partial R(y^n, n)}{\partial n} \right) E^n_\tau g(y^n) \, dn.
\]

**Proof.** See Appendix. \( \square \)

The left-hand side of eq. (31), as well as the first two right-hand side terms, 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. And the second term on the right-hand side reflects the fact that redistribution is most efficiently done late in life when rates of return are increasing with savings. Thus, only the third right-hand side term in eq. (31) is new.

The third term is best understood by considering the tax reform that directly leads to the optimality condition in eq. (31). This tax reform lowers the marginal tax rate on labor income around \( z \), while offsetting any utility gains by raising taxes on capital income. Such a tax reform distorts savings (as captured by the left-hand side of eq. (31)), but might yield an efficiency gain if capital income is increasing in ability (first term on the right-hand side), and if the rich obtain relatively high marginal returns to investment (second term on the right-hand side). At the same time, the increase in taxes on capital income that keeps utility constant must be increasing with ability if for any capital income level \( y^n > y \) individuals’ discount rates are increasing in ability, such that \( \partial R(y^n, n)/\partial n > 0 \). The resulting increase in marginal taxes on capital income for individuals with income above \( y \) generates an additional excess burden \( E^n_\tau \) for those individuals.

Unfortunately, the interpretation of the third right-hand side term in eq. (31) remains rather technical in nature. Moreover, its sign is a priori ambiguous. However, we can obtain some more
insight by taking the derivative of eq. (30):

\[
(32) \quad \frac{\partial R(y^n, n)}{\partial n} = (1 - \tau')(y_{an} - \frac{y_{aa}y_n}{y_a}).
\]

If we combine the two microfoundations of Propositions 2 and 3, it can be shown that this term is strictly negative. That is, if return heterogeneity can be fully attributed to closely-held assets \((y_n \geq 0)\) and scale economies in wealth investment \((y_{aa} \geq 0)\), while the marginal rate of return does not directly depend on ability \((y_{an} = 0)\), then \(\partial R^n / \partial n < 0\). According to eq. (31), this would lead to even higher optimal taxes on capital income. Hence, also in the more general case, Proposition 4 confirms that the combination of these two microfoundations of return heterogeneity yields strictly positive optimal taxes on capital income.

6 Simulations

To be done.

7 Conclusion

To be done.

References


A Proof of Lemma 1

We optimize the tax functions $T$ and $\tau$ to maximize social welfare using tax perturbations. A complication 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. To solve this problem, we follow the Euler-Lagrange formalism. It was first applied to optimal taxation by BohÁ¡cek and Kejak (2005), Werquin et al. (2015), and Lehmann et al. (2018).

We start from a standard proof for the Euler-Lagrange equation (see e.g. Arfken and Weber, 2005, chapter 17). We adapt the standard proof to incorporate behavioral responses to tax reforms. We only prove the optimality for the tax on capital income $\tau$. The proof for the tax on labor income $T$ follows exactly the same steps.

We introduce the tax perturbations in subsection A.1. We also formalize the reform parameters that are used in the main text. Next, in subsection A.2, we study the behavioral responses to the different tax perturbations. We then prove lemma 1 using the Euler-Lagrange approach in subsection A.3, using the results from the preceding subsections.

A.1 Tax reforms

The tax function $\tau$ is optimal if any small perturbation of $\tau$ leaves social welfare unchanged. For any level of capital income $y$, we can introduce a tax reform of size $\epsilon \eta(y)$. The function $\eta$ is an arbitrary, nonlinear but smooth tax reform function. The parameter $\epsilon$ is an infinitesimal, which allows varying the size of the reform. We can construct any small perturbation to $\tau$ by choosing $\eta$ and $\epsilon$. The tax liability at capital income $y$ after the tax reform becomes: $\tau(y) + \epsilon \eta(y)$. If the value of $\epsilon$ is zero, the unreformed tax function $\tau$ is in place. The tax function $\tau$ is optimal if any small perturbation leaves social welfare unchanged. An equivalent way of stating this condition, is that the optimal value of $\epsilon$ is zero for every reform function $\eta$.

We construct the individual budget constraints, taking into account the tax reform. We want to relate the optimality conditions to the behavioral responses in the main text. We thus also take into account reform parameters $\sigma^T$, $\rho^T$, $\sigma^\tau$, $\rho^\tau$. Let $a(y,n)$ be the required level of savings for a type-$n$ individual to get capital income $y$. The individual budget constraints are then as follows:

$$C_1(z,y,n,\sigma^T,\rho^T) \equiv z - a(y,n) - T(z) - z\sigma^T - \rho^T,$$

$$C_2(y,n,\sigma^\tau,\rho^\tau,\epsilon) \equiv a(y,n) + y - \tau(y) - \epsilon \eta(y) - y\sigma^\tau - \rho^\tau.$$

To avoid dense notation, we introduce the following utility function for an individual of type $n$ facing the reformed tax system:

$$U(z,y,n,\sigma^T,\sigma^\tau,\rho^T,\rho^\tau,\epsilon) \equiv u(C_1(z,y,n,\sigma^T,\rho^T), C_2(y,n,\sigma^\tau,\rho^\tau,\epsilon)) - v(z/n).$$

Individuals maximize utility (35) subject to budget constraints (33) and (34). The first-order conditions for the individual optimization problem are as follows:

$$U_z = U_y = 0.$$

The functions $U_z(z,y,n,\sigma^T,\sigma^\tau,\rho^T,\rho^\tau,\epsilon)$ and $U_y(z,y,n,\sigma^T,\sigma^\tau,\rho^T,\rho^\tau,\epsilon)$ correspond to the shift functions introduced by Jacquet et al. (2013).
We denote demand functions for capital and labor income for a type-\( n \) individual for given values of the reform parameters as \( Y(n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon) \) and \( Z(n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon) \). We denote the corresponding indirect utility function as \( V(n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon) \). Apply the envelope theorem to individual objective (35) to find the following property:

\[
\frac{\partial V}{\partial \epsilon} = -u_2 \eta(y).
\]

A.2 Behavioral responses to tax reforms

Suppose there is a marginal change in any parameter \( \nu \). The parameter \( \nu \) could be the type \( n \) of the individual or any of the tax reform parameters. Individuals update their behavior such that their first-order conditions remain satisfied:

\[
0 = \frac{dU_z}{d\nu} = U_{zz} \frac{\partial Z}{\partial \nu} + U_{zy} \frac{\partial Y}{\partial \nu} + U_{z\nu},
\]

\[
0 = \frac{dU_y}{d\nu} = U_{yz} \frac{\partial Z}{\partial \nu} + U_{yy} \frac{\partial Y}{\partial \nu} + U_{y\nu}.
\]

The terms \( \frac{\partial Z}{\partial \nu} \) and \( \frac{\partial Y}{\partial \nu} \) capture the total effects of a marginal change in the parameter \( \nu \) on the taxable incomes. The terms \( \frac{\partial Z}{\partial \nu} \) and \( \frac{\partial Y}{\partial \nu} \) include second-round effects caused by the non-linearity of the tax functions.

If taxable incomes change due to a reform, individuals face new marginal tax rates. These changes in the marginal tax rates have further compensated effects. Jacquet et al. (2013) and Jacobs and Boadway (2014) include similar second-round effects to define behavioral elasticities. Write equations (38)-(39) in matrix notation to find the following lemma. This lemma is a direct application of the general implicit function theorem.

Lemma 1. The effects of a marginal change in any parameter \( \nu \) on labor and capital incomes are given by:

\[
\begin{pmatrix}
\frac{\partial Z}{\partial \nu} \\
\frac{\partial Y}{\partial \nu}
\end{pmatrix} = -\begin{pmatrix}
U_{zz} & U_{zy} \\
U_{yz} & U_{yy}
\end{pmatrix}^{-1}
\begin{pmatrix}
U_{z\nu} \\
U_{y\nu}
\end{pmatrix}.
\]

A change in parameter \( \epsilon \) affects the tax liabilities and marginal tax rates at each income \( y \). A reform to the parameter \( \epsilon \) thus has income and substitution effects on the taxable incomes. The income effects are proportional to \( \eta(y) \) and the substitution effects are proportional to \( \eta'(y) \). We show this formally in the following lemma.

Lemma 2. The behavioral responses to the different tax reforms are related as:

\[
\begin{align*}
\frac{\partial Z}{\partial \epsilon} &= \frac{\partial Z}{\partial \rho^T} \eta(y) + \left( \frac{\partial Z}{\partial \sigma^T} - y \frac{\partial Z}{\partial \rho^T} \right) \eta'(y), \\
\frac{\partial Y}{\partial \epsilon} &= \frac{\partial Y}{\partial \rho^T} \eta(y) + \left( \frac{\partial Y}{\partial \sigma^T} - y \frac{\partial Y}{\partial \rho^T} \right) \eta'(y).
\end{align*}
\]

Proof. The function \( U \) has the following second-order partial derivatives (evaluate for the situation with
no reform, so $\epsilon = \rho^T = \rho^r = \sigma^T = \sigma^r = 0$:

\begin{align}
(43) & \quad U_{\varepsilon \rho^r} = -\frac{v'u_{11}}{nu_1}, \quad U_{\varepsilon \rho^o} = \left( u_{11} - u_{12} \frac{u_1}{w_2} \right) \frac{1}{y_a}, \\
(44) & \quad U_{\varepsilon \sigma^r} = -\frac{v'u_{12}}{nu_1}, \quad U_{\varepsilon \rho^o} = \left( u_{21} - u_{22} \frac{u_1}{w_2} \right) \frac{1}{y_a}, \\
(45) & \quad U_{\varepsilon \sigma^r} = zU_{z\varepsilon \rho^r} - u_1, \quad U_{\varepsilon \sigma^r} = yU_{z\varepsilon \rho^r}, \\
(46) & \quad U_{\varepsilon \sigma^r} = yU_{\varepsilon \rho^o} - u_2, \quad U_{\varepsilon \sigma^r} = zU_{\varepsilon \rho^o}, \\
(47) & \quad U_{\varepsilon \varepsilon} = U_{z\varepsilon \rho^r} \eta(y), \quad U_{\varepsilon \varepsilon} = U_{\varepsilon \rho^o} \eta(y) - u_2 \eta'(y). \\
(48) & \quad U_{zz} = -u_1 T'' + u_{11} \left( \frac{v'}{nu_1} \right)^2 - \frac{v'^2}{n^2}, \\
(49) & \quad U_{z\varepsilon} = -\left( u_{11} - u_{12} \frac{u_1}{w_2} \right) \frac{1}{y_a} \frac{v'}{nu_1}, \\
(50) & \quad U_{yy} = \left[ u_{11} - 2u_{12} \frac{u_1}{w_2} + u_{22} \left( \frac{u_1}{w_2} \right)^2 - (u_2 - u_1) \frac{y_{aa}}{y_a} \right] \frac{1}{y_a^2} - u_2 T''', \\
(51) & \quad U_{zn} = \frac{v'}{n} \left( \frac{u_{11} - u_{21} y_n}{u_1} + \frac{1}{n} \left( 1 + \frac{v'^2}{v'} \right) \right), \\
(52) & \quad U_{yn} = -\frac{y_n}{y_a} \left[ (u_{11} - u_{12}) - (u_{21} - u_{22}) \frac{u_1}{w_2} + (u_2 - u_1) \left( \frac{y_{aa}}{y_n} - \frac{y_{aa}}{y_a} \right) \right].
\end{align}

Verify the following relation between equations (43)-(52):

\begin{align}
(53) & \quad U_{\varepsilon \varepsilon} = U_{z\varepsilon \rho^r} \eta(y) + (U_{z\varepsilon \sigma^r} - yU_{z\varepsilon \rho^r}) \eta'(y), \\
(54) & \quad U_{\varepsilon \varepsilon} = U_{\varepsilon \rho^o} \eta(y) + (U_{\varepsilon \sigma^r} - yU_{\varepsilon \rho^o}) \eta'(y).
\end{align}

Substitute equation (40) for the partial derivatives of $U$ to find equations (41) and (42). \hfill \Box

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

Consider the optimization problem of a government choosing the value of $\epsilon$, for a given reform function $\eta$. The Lagrangian for this maximization problem is given by:

\begin{align}
\Lambda(\epsilon) \equiv \int_0^\infty \mathcal{V}(\epsilon, n)f(n)dn + \lambda_1 \int_0^\infty [T(Z(\epsilon, n)) - g_1]f(n)dn \\
+ \lambda_2 \int_0^\infty [\tau(\mathcal{V}(\epsilon, n)) + \epsilon \eta(\mathcal{V}(\epsilon, n)) - g_2]f(n)dn.
\end{align}

Here, we used short-hand notations for the function arguments, ignoring the other reform parameters. We assume that this objective function is sufficiently smooth, excluding kinks and bunching.

The optimal value of $\epsilon$ is characterized by the first-order condition $\partial \Lambda/\partial \epsilon = 0$. We want the optimal value of $\epsilon$ to be zero for every reform function $\eta$. We thus evaluate the optimality condition for $\epsilon = 0$:

\begin{align}
\frac{\partial \Lambda(0)}{\partial \epsilon} = 0 \Rightarrow 0 = \int_0^\infty \left( \frac{1}{\lambda_1/\lambda_2} - \frac{u_2}{\lambda_1} + \frac{\partial Z}{\partial \rho^r} T' + \frac{\partial \mathcal{V}}{\partial \rho^o} \frac{T'}{\lambda_1/\lambda_2} \right) g(y) \eta(y)dy \\
+ \int_0^\infty \left[ \left( \frac{T'}{T} g(y) \right) \eta'(y)dy \\
+ \int_0^\infty \left[ \left( \frac{\partial \mathcal{V}}{\partial \sigma^r} - \frac{\partial Z}{\partial \rho^o} \frac{T'}{\lambda_1/\lambda_2} g(y) \right) \eta'(y)dy. \right]
\end{align}
Here we used property (37) and lemma 2. We divided the condition by $\lambda_1$. We also changed the integration variables. To do so, we used the identity $dF(n) = dG(y^n) \Leftrightarrow f(n) \, dn = g(y^n) \, dy^n$. The latter identity follows from the monotonicity of the allocation. The term $y^n$ indicates the value of capital income $y$ for an individual of type $n$.

We now relate equation (56) to the notations in the main text. Note the following identities:

\begin{align*}
\frac{d\tilde{z}}{d\rho^T} &= \frac{\partial Z}{\partial \rho^T}, \\
\frac{d\tilde{z}}{d\sigma^T} &= \frac{\partial Z}{\partial \sigma^T}, \\
\frac{d\tilde{z}}{d\rho^T} &= \frac{\partial Z}{\partial \sigma^T}. \\
\end{align*}

Substitute these identities into (56), and perform partial integration on the second and third lines:

\begin{equation}
0 = \int_0^{\infty} \left[ \left( \frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) + \frac{dE_T(y)g(y)}{dy} \right] \eta(y) \, dy \\
+ E_T(\infty)g(\infty)\eta(\infty) - E_T(0)g(0)\eta(0).
\end{equation}

Here we use definitions $\alpha_2$ and $E_T$ from the main text.

This condition must hold for every reform function $\eta$. Suppose that the term within square brackets differs from zero on some interval. Choose $\eta$ such that it is zero at the endpoints. The second line of (59) then becomes zero. Furthermore, let $\eta$ have the same sign everywhere as the term within square brackets. It follows that the value of the integral is strictly positive. Condition (59) however tells us that the value of the integral must be zero. It follows by contradiction that the terms between the square brackets must sum to zero for every capital income $y$. The latter result is an application of the fundamental theorem of the calculus of variations. This yields the Euler-Lagrange equation for the optimal tax on capital income:

\begin{equation}
\forall y : \left( \frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) = -\frac{d}{dy} \left[ E_T(y)g(y) \right].
\end{equation}

We thus find that the first line of (59) must be zero. It follows that the second line of (59) must also be zero for every reform function $\eta$. This yields the corresponding transversality conditions:

\begin{equation}
E_T(0)g(0) = 0 \text{ and } E_T(\infty)g(\infty) \to 0.
\end{equation}

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

\begin{equation}
\forall z : (1 - \alpha_1) h(z) = -\frac{d}{dz} \left[ E_T(z)h(z) \right],
\end{equation}

with transversality conditions:

\begin{equation}
E_T(0)h(0) = 0 \text{ and } E_T(\infty)h(\infty) \to 0.
\end{equation}

Transversality conditions (61) and (63) form a system of equations in the marginal tax rates at the end points. Solving it yields marginal tax rates at the bottom: $T'(0) = \tau'(0)$.\footnote{We assumed that the government objective function is sufficiently smooth. This assumption precludes discontinuous jumps in the density functions. If on the contrary there were a mass point at the bottom, then the optimal marginal tax rate at the bottom would no longer be zero.}
B Proof of Propositions 1, 2 and 3

Propositions 1 and 2 follow as special cases from Proposition 3. To prove Proposition 3, we will substitute the optimal tax on labor income into the optimal tax on capital income. This yields a condition for the optimal mix of taxes on labor income and capital income. To interpret this condition, we need to relate the effects of reforms to both tax functions. In subsection B.1 we derive the Slutsky symmetry between the tax bases. It relates effects of capital taxes on labor income to effects of labor taxes on capital income. Next, in subsection B.2, we relate income effects in the two periods. This allows rewriting the optimal capital tax in terms of first-period welfare weights, in subsection B.3. Finally, in subsection B.4, we use the results of the preceding steps to prove proposition 3.

B.1 Slutsky symmetry

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

**Lemma 3.** Cross-prices responses of labor income and capital income comply to the following Slutsky symmetry:

\[(64) \frac{d\tilde{z}}{d\sigma^\tau} - y \frac{d\tilde{z}}{d\rho^\tau} = \frac{1}{R} \left( \frac{d\tilde{y}}{d\sigma^T} - z \frac{d\tilde{y}}{d\rho^T} \right).\]

**Proof.** We first construct compensated reforms to the marginal tax rates. Denote the taxable incomes in the situation with before any reform as \(z^0 \equiv Z(n, 0, 0, 0, 0, 0)\) and \(y^0 \equiv Y(n, 0, 0, 0, 0, 0)\). To determine compensated responses to an increase in marginal tax rates, we fix \(\rho^T = -\sigma^T z^0\) and \(\rho^\tau = -\sigma^\tau y^0\).

We first show that these reforms are indeed compensated. Apply the envelope theorem to objective function (35). This yields the following properties:

\[(65) \frac{\partial V}{\partial \sigma^T} \bigg|_{\rho^T = -\sigma^T z^0} = \frac{\partial V}{\partial \sigma^T} - \frac{\partial V}{\partial \rho^T} z^0 = -[Z(n, \sigma^T, \sigma^\tau, -\sigma^T z^0, -\sigma^\tau y^0, 0) - z^0] u_1,\]

\[(66) \frac{\partial V}{\partial \sigma^\tau} \bigg|_{\rho^\tau = -\sigma^\tau y^0} = \frac{\partial V}{\partial \sigma^\tau} - \frac{\partial V}{\partial \rho^\tau} y^0 = -[Y(n, \sigma^T, \sigma^\tau, -\sigma^T z^0, -\sigma^\tau y^0, 0) - y^0] u_2.\]

The last two expressions are zero in the situation without reforms, when \(\rho^T = \rho^\tau = \sigma^T = \sigma^\tau = 0\). We thus indeed constructed compensated reforms to the marginal tax rates.

Evaluate the second-order partial derivatives for the situation with no reform:

\[(67) \frac{\partial^2 V}{\partial \sigma^T \partial \sigma^T} \bigg|_{\rho^T = -\sigma^T z^0} = - \left( \frac{d\tilde{z}}{d\sigma^T} - z^0 \frac{d\tilde{z}}{d\rho^T} \right) u_1,\]

\[(68) \frac{\partial^2 V}{\partial \sigma^T \partial \sigma^\tau} \bigg|_{\rho^\tau = -\sigma^\tau y^0} = - \left( \frac{d\tilde{y}}{d\sigma^T} - y^0 \frac{d\tilde{y}}{d\rho^T} \right) u_2.\]

Here we used identities (57)–(58) to write our result in the notations of the main text. **Young’s theorem** now imposes that the second-order derivatives of the indirect utility function are equal. Apply this requirement to (67) and (68) to find Slutsky symmetry (64). \(\Box\)
B.2 Relating income effects in periods 1 and 2

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 in the second period increases instead, individuals decrease their savings. Furthermore, both responses have different effects on the marginal tax on capital income. The reason is that we allow taxes on capital income to be nonlinear. The different effects on the marginal tax rates 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 4. (1) The effects of changes to unearned incomes in the first and the second period are related as follows:

\[
\begin{align*}
R \frac{\partial \tilde{z}}{\partial \rho^T} &= \frac{\partial \tilde{z}}{\partial \rho^T} - \frac{\partial \tilde{z}}{\partial \sigma^T} \frac{\partial R}{\partial y}, \\
R \frac{\partial \tilde{y}}{\partial \rho^T} &= \frac{\partial \tilde{y}}{\partial \rho^T} + y_a - \frac{\partial \tilde{y}}{\partial \sigma^T} \frac{\partial R}{\partial y}.
\end{align*}
\]

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

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

\[
\alpha_2 R = \alpha_1 - \frac{\partial R}{\partial y} E_\tau - \frac{\tau'}{\lambda_1/\lambda_2} y_a.
\]

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

\[
\frac{\partial R}{\partial y} = -\tau'' y_a + \left( \frac{u_1}{u_2} - 1 \right) \frac{y_{1a}}{y_a^2}.
\]

Use (72) to verify the following relation between equations (43)–(52):

\[
\begin{align*}
\frac{\partial R}{\partial y} &= -\tau'' y_a + \left( \frac{u_1}{u_2} - 1 \right) \frac{y_{1a}}{y_a^2}.
\end{align*}
\]

B.3 Optimal tax on capital income in terms of first-period social welfare weights

Using lemma 4 above allows rewriting the optimal capital income tax in terms of first-period social welfare weights. We show this in the following proposition.
Lemma 5. The optimal capital income tax can be written as follows:

$$RE_\tau g(y^n) = \int_{y^n}^{\infty} \left[ 1 - \alpha^1 + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} \left( \frac{dR}{dy} - \frac{\partial R}{\partial y} E_\tau \right) \right] g(y)dy.$$  \hspace{1cm} (76)

Proof. (a) The first fundamental theorem of calculus and transversality condition (61) yield the following expression:

$$RE_\tau g(y^n) = - \int_{y^n}^{\infty} \frac{d}{dy} (RE_\tau g(y))dy$$  \hspace{1cm} (77)

(b) Substitute (71) into the optimal capital income tax and use the individuals’ Euler equation to find:

$$E_\tau g(y^n) = \int_{y^n}^{\infty} \left[ \frac{1}{\lambda_1/\lambda_2} - \frac{1}{R} \left( \alpha^1 - \frac{\partial R}{\partial y} E_\tau - \frac{\tau'}{\lambda_1/\lambda_2} y_a \right) \right] g(y)dy$$

$$= \int_{y^n}^{\infty} \left[ \frac{1}{R} \left( 1 + \frac{R}{\lambda_1/\lambda_2} - 1 - \alpha^1 + \frac{\partial R}{\partial y} E_\tau + \frac{\tau' y_a}{\lambda_1/\lambda_2} \right) \right] g(y)dy$$

$$= \int_{y^n}^{\infty} \frac{1}{R} \left( 1 - \alpha^1 + \frac{\partial R}{\partial y} E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} \right) g(y)dy.$$  \hspace{1cm} (78)

Take the derivatives with respect to $y$ on both sides:

$$\frac{d}{dy} (E_\tau g(y)) = - \frac{1}{R} \left( 1 - \alpha^1 + \frac{\partial R}{\partial y} E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} \right) g(y).$$  \hspace{1cm} (80)

Substitute the latter into (77) to prove the lemma.

B.4 Proof of Proposition 3

Rewrite optimal labor tax by bringing the marginal tax on labor income to one side:

$$T' = - \frac{1}{g(y)} \left( \frac{d\tilde{z}}{d\sigma^r} - y \frac{d\tilde{z}}{d\rho^r} \right)^{-1} \int_{y^n}^{\infty} \left( \frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y)dy$$

$$- \frac{\tau'}{\lambda_1/\lambda_2} \left( \frac{d\tilde{z}}{d\sigma^r} - y \frac{d\tilde{z}}{d\rho^r} \right)^{-1} \left( \frac{\bar{y}}{\lambda_1/\lambda_2} \right) \cdot$$

Substitute this value of $T'$ into the optimal capital tax (76) and use the identities $f(n) = h(z)d\tilde{z}/dn = g(y)dy/dn$:

$$\frac{\tau'}{1 + r} \left[ \left( \frac{d\bar{y}}{dn} - \frac{d\bar{y}}{du_1} \right) \left( \frac{d\tilde{z}}{d\sigma^T} - y \frac{d\tilde{z}}{d\rho^T} \right) \left( \frac{d\tilde{z}}{d\sigma^T} - z \frac{d\tilde{z}}{d\rho^T} \right)^{-1} \left( \frac{d\tilde{z}}{d\sigma^T} - y \frac{d\tilde{z}}{d\rho^T} \right) \right]$$

$$= \left[ \frac{d\tilde{z}}{dn} \left( \frac{d\tilde{z}}{d\sigma^T} - z \frac{d\tilde{z}}{d\rho^T} \right)^{-1} \left( \frac{d\tilde{z}}{d\sigma^T} - y \frac{d\tilde{z}}{d\rho^T} \right) - \frac{d\bar{y}}{du_1} \right] \frac{1}{f(n)} \int_{n}^{\infty} (1 - \alpha^1) f(\tilde{n}) d\tilde{n}$$

$$- \frac{d\bar{y}}{du_1} \frac{1}{f(n)} \int_{y}^{\infty} \left[ - \left( \frac{dR}{dy} - \frac{\partial R}{\partial y} \right) E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} \right] g(\bar{y})dy.$$  \hspace{1cm} (82)
To interpret this equation, note first the following identities:

\[
\frac{d\tilde{y}}{dn} = \frac{\partial \tilde{y}}{\partial z} \frac{d\tilde{z}}{dn} + \frac{\partial \tilde{y}}{\partial n},
\]

\[
\frac{d\tilde{y}}{d\sigma^\tau} - y \frac{d\tilde{y}}{d\rho^\tau} = \left( \frac{\partial \tilde{y}}{\partial \sigma^\tau} - y \frac{\partial \tilde{y}}{\partial \rho^\tau} \right) + \frac{\partial \tilde{y}}{\partial z} \left( \frac{d\tilde{z}}{d\sigma^\tau} - y \frac{d\tilde{z}}{d\rho^\tau} \right),
\]

\[
\frac{d\tilde{y}}{d\sigma^T} - z \frac{d\tilde{y}}{d\rho^T} = \frac{\partial \tilde{y}}{\partial z} \left( \frac{d\tilde{z}}{d\sigma^T} - z \frac{d\tilde{z}}{d\rho^T} \right).
\]

We used the separability of preferences between consumption and leisure in the last equation. Substitute these three identities into (82) and use Slutsky symmetry (64) to rewrite the optimal tax equation:

\[
\tau'y_a \left[ R \frac{1}{y} \frac{\partial \tilde{y}}{\partial \sigma^\tau} - y \frac{\partial \tilde{y}}{\partial \rho^\tau} \right] = -\left( \frac{n}{y} \frac{\partial \tilde{y}}{\partial n} \right) \left( \frac{n}{z} \frac{d\tilde{z}}{dn} \right)^{-1} \frac{1}{zh(z)} \int_{z}^{\infty} (1 - \alpha^1) h(\tilde{z}) d\tilde{z} - \frac{1}{yg(y)} \int_{y}^{\infty} \frac{1 + \tilde{y} - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} g(\tilde{y}) d\tilde{y} + \frac{1}{yg(y)} \int_{y}^{\infty} \left( \frac{dR}{dy} - \frac{\partial R}{\partial y} \right) E_\tau g(\tilde{y}) d\tilde{y}.
\]

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 \( \xi_{y|z}^* \), \( \xi_{z}^* \), \( e_{y|z}^* \) and \( e_{z}^* \) and the optimal labor tax to find Proposition 3.