THE ECONOMIC JOURNAL

The Economic Journal, **135** (*January*), 180–211 https://doi.org/10.1093/ej/ueae083
Advance Access Publication Date: 16 September 2024

OPTIMAL TAXATION OF CAPITAL INCOME WITH HETEROGENEOUS RATES OF RETURN*

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

We derive the Pareto-efficient mix of non-linear taxes on labour income and capital income if people differ in their rates of return on capital. We allow for two reasons why rates of return differ: because individuals with higher ability are better able to invest their capital or because wealthier individuals enjoy scale effects in wealth accumulation. In both cases, a strictly positive tax on capital income is part of any Pareto-efficient tax system. We derive a condition for the Pareto-efficient tax mix that relies solely on empirical sufficient statistics—not on social welfare weights—and find that Pareto-efficient taxes on capital income increase with the degree of return heterogeneity. Numerical simulations for empirically plausible return heterogeneity suggest that Pareto-efficient marginal tax rates on capital income are positive and substantial.

https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-10.html
How should the burden of taxation be distributed between labour and capital? We aim to answer this question, while taking into account two recent empirical findings on wealth and its returns. First, net wealth is primarily composed of previously earned labour income, which suggests a close link between labour income and capital income.[1] Second, a growing literature documents significant differences in the rates of return that individuals earn on their wealth. In particular, people tend to obtain higher returns if they are more able investors (‘*type-dependent returns*’) and if they have more wealth to invest (‘*scale-dependent returns*’).[2]

The first major insight of our paper is that the optimal *mix* of taxes on labour and capital income depends solely on efficiency considerations, not on the government’s distributional preferences.[3] Intuitively, we capture the close link between labour income and capital income through a one-to-one correspondence between individual ability and both types of income. As a result, adjustments to either labour-income taxes or capital-income taxes can achieve the same distributional effects.

* Corresponding author: Aart Gerritsen, Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Email: agerritsen@ese.eur.nl.

This paper was received on 23 April 2021 and accepted on 12 September 2024. The Editor was Francesco Lippi.

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.13683258.

Any opinions and arguments expressed in this paper are personal and do not necessarily reflect the official views of the OECD or its Member Countries. We are grateful to Francesco Lippi and three anonymous referees for their insightful and constructive remarks. We also thank Spencer Bastani, Bob Bremer, Giacomo Corneo, Katherine Cuff, André Decoster, Emanuel Hansen, Albert Jan Hummel, Tilman Klumpp, Tuomas Kosonen, Claus Kreiner, Etienne Lehmann, Agnar Sandmo, Floris Zoutman and participants of numerous seminars and conferences for helpful comments and discussions. Bas Jacobs acknowledges support from the Netherlands Institute for Advanced Studies (Duisenberg Fellowship 2018–2019). Kevin Spiritus acknowledges the financial support of BELSPO via BRAIN.be project BR/121/A5/CRESUS.

[1] See, e.g., Black *et al.* (2023) on Norway and Kaymak *et al.* (2022) on the United States.
[2] Key recent papers are Fagereng *et al.* (2020) on Norway and Bach *et al.* (2020) on Sweden; the next section provides a more elaborate review of the empirical literature on return heterogeneity.
[3] To be more precise, the optimal distribution of tax burdens across individuals *does* depend on distributional preferences. But the optimal decomposition of burdens into taxes on labour income and taxes on capital income *does not* depend on distributional preferences.

The optimal mix of both taxes then minimises the efficiency costs associated with a *given* distribution of resources.

The second major insight of our paper is that capital income should be taxed alongside labour income as long as individuals differ in their rates of return. This is true whether return heterogeneity originates from type or scale dependence. Intuitively, type-dependent returns imply that capital income contains ability rents. Scale-dependent returns imply a market failure, as it would be mutually beneficial if the wealthy were to invest on behalf of the poor. As a result, taxes on capital income are more efficient because they tax rents (in case of type-dependent returns) or because they alleviate a market failure (in case of scale-dependent returns). This ensures that a tax on capital income is part of the policy optimum.

Because the optimal mix of taxes on labour and capital income is independent of distributional preferences, we refer to it as the *Pareto-efficient* tax mix. We derive an expression for the Pareto-efficient tax mix that only depends on a small number of empirical sufficient statistics, without reference to normatively ambiguous social welfare weights. The most important sufficient statistic measures the extent to which rates of return differ across individuals. The Pareto-efficient tax mix features higher taxes on capital income if there is more return heterogeneity. It furthermore depends in intuitive ways on the compensated elasticities of both capital income and labour income with respect to their tax rates and on the hazard rate of the capital income distribution. Numerical simulations for the United States suggest that the Pareto-efficient tax mix features marginal tax rates on capital income that are positive and substantial.

Our study builds on a large literature that argues in favour of taxes on capital income.[4] Most closely related are a few papers that also study optimal capital taxes with return heterogeneity (Gahvari and Micheletto, 2016; Kristjánsson, 2016; Guvenen *et al.*, 2023). Compared to these papers, our main contributions are two-fold. We derive conditions for the *Pareto-efficient mix* of taxes on capital and labour income. And we show that both type- and scale-dependent returns yield positive Pareto-efficient taxes on capital income.

We derive our results within a deterministic two-period version of the Mirrlees (1971) framework. We abstract from risk and focus our analysis solely on the implications of non-random differences in rates of return. We do this because risk cannot fully explain observed differences in rates of return (e.g., Fagereng *et al.*, 2020). Individuals differ in a single exogenous characteristic, namely their ability. In the first period, they choose how much to work and how much to save. Labour income equals the product of labour supply and ability. In the second period, individuals consume their savings plus the capital income they earn with their savings. The main innovation of our model is to define capital income as a general and possibly non-linear function of both savings and ability. This allows us to capture both type- and scale-dependent capital returns and derive their distinct implications for the Pareto-efficient tax mix.

We first consider *type-dependent* returns. In this case, rates of return are increasing in ability. This may be due to a positive association between ability and entrepreneurial talents. Capital income then reflects both savings—which were previously earned as labour income—and ability rents. As is well known, a pure rent tax is non-distortionary. As a result, compared to a tax on labour income, a tax on capital income can achieve the same distributional effects with less distortions in labour supply. At the same time, unlike a tax on labour income, a tax on capital income does distort savings decisions. The Pareto-efficient tax mix trades off the benefits of reduced labour-supply distortions against the costs of larger saving distortions. The larger the

[4] The next section provides an overview of this literature. More thorough reviews are provided by Diamond and Saez (2011), Jacobs (2013) and Bastani and Waldenström (2020).

degree of return heterogeneity, the more important ability rents in capital income and thus the larger the Pareto-efficient tax on capital income.

We then consider *scale-dependent* returns. In this case, rates of return are increasing in the level of wealth, but do not directly depend on ability. This may reflect fixed costs of wealth-management services that only sufficiently wealthy people can afford to pay. As a result, wealthy people ('the rich') earn a higher marginal rate of return on their capital than people with little wealth ('the poor'). It would be mutually beneficial if the rich were to save on behalf of the poor, but an implicit market failure prevents such transactions from taking place. We show that a positive tax on capital income helps to alleviate this market failure. Specifically, the government could reduce marginal taxes on labour income and raise marginal taxes on capital income. Such a policy transfers funds from the poor to the rich in the first period and from the rich back to the poor in the second period—implicitly forcing the rich to save on behalf of the poor. The Pareto-efficient mix of both taxes trades off the benefits of alleviating the market failure against the costs of savings distortions. As before, the Pareto-efficient tax on capital income is positive and increasing in the extent of return heterogeneity.

Besides deriving analytical results, we numerically simulate our model to obtain a quantitative sense of the importance of return heterogeneity for the Pareto-efficient tax mix. We do so under the assumption that return heterogeneity is solely driven by either type- or scale-dependent returns. We calibrate our model on the basis of US data on the distribution of income, but model return heterogeneity by using Norwegian estimates from Fagereng *et al.* (2020). In our simulations, we first impose the actual US tax schedule and then adjust the mix of taxes until no further Pareto improvements are feasible. The resulting Pareto-efficient tax mix yields the same utility for each taxpayer as the actual US tax system, but it generates fewer distortions and thus more government revenue.

The simulated Pareto-efficient tax mix features positive and substantial taxes on capital income regardless of the source of return heterogeneity. In our baseline simulations, Pareto-efficient tax rates on capital income are on average around 10% in case of type-dependent returns and around 25% in case of scale-dependent returns. The Pareto-efficient tax rate at the highest income decile is around 17% for both type- and scale-dependent returns. There are important differences between type- and scale-dependent returns when it comes to the shape of the non-linear tax schedules. With type-dependent returns, Pareto-efficient tax rates on capital income are monotonously increasing in income and lower than marginal taxes on labour income for every individual. In contrast, with scale-dependent returns, Pareto-efficient tax rates on capital income are hump-shaped and exceed tax rates on labour income for a majority of individuals. These results highlight the importance of further empirical research on why rates of return differ across individuals.

Our results are derived within a stylised model that necessarily abstracts from aspects of reality. In particular, we do not account for various other potential reasons for positive taxes on capital income, such as heterogeneous preferences or political constraints. Nor do we allow for true multidimensional heterogeneity, which would break the one-to-one correspondence between labour income and capital income. Nevertheless, we show that a relatively simple model of one-dimensional heterogeneity and differences in rates of return is sufficient to justify significant tax rates on capital income.

The remainder of the paper is organised as follows. In Section 1, we first briefly discuss the empirical evidence on return heterogeneity. We then provide an elaborate discussion of earlier results on optimal taxation of capital income and indicate how we contribute to this large literature. In Section 2, we introduce and discuss the theoretical setting of our paper. In Section 3, we explicitly show how our model is able to capture two plausible micro-foundations of return heterogeneity. In Section 4, we derive and discuss expressions for the Pareto-efficient tax mix in terms of sufficient statistics. Section 5 provides numerical simulations of Pareto-efficient taxes on labour and capital income. A final section concludes.

1. Related Literature

1.1. *Empirical Evidence on Return Heterogeneity*

Our research is motivated by a large and growing number of studies documenting the empirical importance of return heterogeneity. First, there is direct evidence on return heterogeneity. The seminal paper by Yitzhaki (1987) studies a subset of US tax returns from 1973 and finds that rates of return increase with income. Piketty (2014) and Saez and Zucman (2016) show that rates of return on the endowments of US universities and other foundations are increasing in the size of the endowments. More recently, Bach *et al.* (2020) and Fagereng *et al.* (2020) study administrative data on the populations of Norway and Sweden over several years and find convincing evidence of return heterogeneity. For example, moving from the 10th to the 90th percentile of financial wealth, Fagereng *et al.* (2020) find that the average rate of return increases by 1.6 percentage points. This figure is only slightly lower if they restrict attention to safe assets or if they control for risks in underlying portfolios. Bastani *et al.* (2023) study Swedish data and find that returns to capital are increasing with measures of cognitive ability.

Second, a large literature in finance provides evidence that richer individuals tend to make fewer mistakes in their investments. See Campbell (2016) for an overview. An abundance of evidence shows that individuals do not optimally diversify their portfolios (e.g., Benartzi and Thaler, 2001; Choi *et al.*, 2005; Calvet *et al.*, 2007; Goetzmann and Kumar, 2008; Von Gaudecker, 2015). Individuals consistently fail to optimise their financial portfolio even conditional on risk, for example, by exposing themselves to excess interest and fee payments (Barber *et al.*, 2005; Agarwal *et al.*, 2009; Choi *et al.*, 2010; 2011). Unsurprisingly, investment mistakes are linked to individuals' financial literacy or sophistication, which is itself positively associated with education and wealth (e.g., Lusardi and Mitchell, 2011; Van Rooij *et al.*, 2011; Lusardi *et al.*, 2017). A natural implication of this evidence is that richer individuals obtain higher rates of return on their savings. Indeed, Lusardi *et al.* (2017) suggest that 30–40% of inequality in US retirement wealth can be attributed to differences in financial sophistication.

Third, recent simulations suggest that return heterogeneity is necessary to reconcile life-cycle models with observed patterns of wealth inequality. In particular, Gabaix *et al.* (2016) and Benhabib *et al.* (2019) argue that return heterogeneity is needed to explain the dynamics of the fat, right tail of the US wealth distribution. Importantly, Gabaix *et al.* (2016) emphasise both 'type dependence' and 'scale dependence' in return heterogeneity. That is, they argue that rates of return could depend on both the underlying individual type—e.g., cognitive ability—and the level of individual wealth. We make explicit use of this distinction in our own model.

1.2. *Literature on Optimal Capital Taxation*

Arguments against taxing capital income date back to at least Mill (1848) and Pigou (1928). They argued that a tax on capital income amounts to taxing labour income twice: first when it is earned, then when it is saved. Modern incarnations of this argument can be found in Atkinson and Stiglitz (1976), Judd (1985) and Chamley (1986). In Atkinson and Stiglitz (1976), taxes on capital income generate the same redistribution and labour-supply distortions as taxes on labour income, but they additionally distort savings. As a result, it would be better not to tax capital income at all. Judd (1985) and Chamley (1986) have shown that taxes on capital income are zero in the steady state of a representative-agent Ramsey model of optimal taxation without any distributional concerns.[5] Much of the subsequent literature explores conditions under which capital taxes are optimal after all. Surveys of this literature can be found in Diamond and Saez (2011), Jacobs (2013) and Bastani and Waldenström (2020).

Taxes on capital may be optimal because savings are relatively complementary to leisure (Corlett and Hague, 1953; Atkinson and Stiglitz, 1976; Erosa and Gervais, 2002; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018); because of tax base shifting between labour or entrepreneurial income and capital income (Christiansen and Tuomala, 2008; Reis, 2010); because physical capital is a substitute for human capital (Jacobs and Bovenberg, 2010); because of heterogeneous preferences for wealth itself (Saez and Stantcheva, 2018); because inheritances positively correlate with labour income (Cremer *et al.*, 2001); because of dynamic inefficiencies in capital accumulation across overlapping generations (Ordover and Phelps, 1979; Atkinson and Sandmo, 1980; King, 1980); because of political constraints and lack of commitment (Farhi *et al.*, 2012; Scheuer and Wolitzky, 2016); because of borrowing constraints (Hubbard and Judd, 1986; Aiyagari, 1995); because of missing insurance markets and idiosyncratic risk in labour productivity (Diamond and Mirrlees, 1978; Golosov *et al.*, 2003; Conesa *et al.*, 2009; Jacobs and Schindler, 2012); and because of uncertainty in capital returns (Varian, 1980; Gordon, 1985; Christiansen, 1993; Schindler, 2008; Shourideh, 2014; Boadway and Spiritus, 2024).

Closer to our paper, it has been shown that capital income should be taxed if preferences to save are positively correlated with ability (Mirrlees, 1976; Saez, 2002; Diamond and Spinnewijn, 2011; Golosov *et al.*, 2013; Ferey *et al.*, 2024; Hellwig and Werquin, 2024). Capital income is then optimally taxed because it is driven by ability, as well as labour income. Our model with type-dependent returns is mathematically isomorphic to a model with heterogeneous preferences to save. Hence, we also find that capital income should be taxed in the presence of type-dependent returns. Naturally, even if models with type-dependent returns or heterogeneous preferences are mathematically equivalent, heterogeneous returns and heterogeneous saving preferences are empirically distinct concepts. We therefore consider the case of type-dependent returns as complementary to the literature on optimal capital taxation and heterogeneous preferences. Furthermore, our model with scale-dependent returns is *not* isomorphic to models with heterogeneous preferences to save. In that case, the rationale for positive taxes on capital income originates from a market failure that is absent from papers on heterogeneous preferences.[6]

Our paper is most closely related to a few papers that also study optimal taxation with heterogeneous returns to capital. Stiglitz (1985; 2000; 2018) conjectures, but does not formally show, that optimal taxes on capital income are positive if rates of return depend on ability. We confirm this conjecture. Gahvari and Micheletto (2016) and Kristjánsson (2016) study the two-type optimal tax framework of Stiglitz (1982) and show that optimal taxes on capital income are

[5] It is often argued that positive taxes on capital income are undesirable as they would imply exponentially growing inter-temporal distortions in consumption that are inconsistent with Ramsey principles, see, for example, Banks and Diamond (2010). Jacobs and Rusu (2018) argue instead that in the steady state of the Chamley–Judd model, taxes on capital income cannot alleviate the distortions of taxes on labour income. Thereby, the zero capital tax in the steady state comes down to a vanishing Corlett and Hague (1953) motive for using capital taxes.

[6] Online Appendix D formally proves that our model with type-dependent returns is mathematically isomorphic to a model with heterogeneous preferences to save, while our model with scale-dependent returns is not.

positive if rates of return are higher for the high-ability type. We contribute to these papers in a number of ways. First, we derive conditions for the Pareto-efficient mix of taxes on capital income and labour income, which do not depend on distributional preferences. Second, we show that Pareto-efficient taxes on capital income are also positive if rates of return are scale dependent.[7] Third, we study an economy with a continuum of types, as in Mirrlees (1971). This allows us to derive meaningful optimal tax formulas in terms of sufficient statistics, as well as gain more insight into the shape of the Pareto-efficient non-linear tax schedule on capital income. Fourth, we provide numerical simulations of Pareto-efficient non-linear taxes on capital income and show that they are positive and substantial.[8]

Guvenen *et al.* (2023) separately consider optimal linear taxation of wealth and capital income in a quantitative overlapping-generations model with heterogeneous returns that originate from binding borrowing constraints. The government wants to reallocate capital from ‘inefficient’ investors with a low rate of return to ‘efficient’ investors with a high rate of return. In their setting, this can be achieved better with a wealth tax than with a tax on capital income. In fact, their optimal tax on capital income is negative. Our mechanisms for optimal positive taxes on capital income are different and we view both papers as complementary. We furthermore differ by linking our results to type- and scale-dependent returns, by deriving conditions for the Pareto-efficient tax mix, which are written in terms of sufficient statistics, and by considering non-linear tax schedules on capital income.

2. Model

2.1. *Individual Behaviour*

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

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

We allow for a general capital-income function $y^n \equiv y(a^n, n)$, which gives capital income as a function of the level of savings and individual ability. As we show later, this capital-income function allows us to capture plausible micro-foundations of return heterogeneity related to closely held businesses and scale economies in wealth accumulation. The case where returns on the assets from a closely held business are increasing in the owner’s ability could be captured by $y_n > 0$. Increasing rates of return in total wealth of an individual could be captured by $y_a > 0$

[7] We show that scale dependence allows the government to improve the allocation of capital by reducing first-period taxes on labour income and raising second-period taxes on capital income. In contrast, Gahvari and Micheletto (2016) and Kristjánsson (2016) conclude that scale dependence of returns does *not* provide a reason to tax capital income. But this conclusion is driven by their assumption that all taxes are levied in the same period.

[8] A recent paper by Schulz (2023) compares optimal capital taxes with type-dependent returns to optimal capital taxes with scale-dependent returns. His main finding is that scale-dependent returns magnify the elasticity of savings: a reduction in savings leads to a reduction in returns, in turn leading to a further reduction in savings. This does not affect optimal tax rules that are—like ours—written in terms of sufficient statistics. An important difference is that we focus on how taxes on capital income can optimally supplement labour-income taxes, while his paper mostly abstracts from optimal labour-income taxes.

and $y_{aa} > 0$. In the latter case, individuals differ in their marginal rate of return y_a. Thus, we implicitly allow for capital-market failures, such that differences in marginal rates of return are not necessarily arbitraged away. Taxes on capital income are denoted by τ^n and second-period consumption equals the sum of savings and after-tax capital income:

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

T^n is a non-linear tax function of labour income z^n and τ^n is a non-linear tax function of capital income y^n. We parameterise the tax schedules in a way that allows us to study the effects of exogenous shifts in their slopes and intercepts. This also allows us to define behavioural elasticities.[9] We write the tax schedules as the following functions:

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

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

where ρ^T in (3) and ρ^τ in (4) are parameters that shift the intercepts of the tax schedules, while σ^T and σ^τ are parameters that shift the slopes of the tax schedules. This parameterisation does not impose any restrictions on the tax schedules because $\tilde{T}(z^n)$ and $\tilde{\tau}(y^n)$ are fully non-linear functions of the tax base.

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

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

Utility of consumption $u(\cdot)$ is increasing, concave and three times continuously differentiable. Disutility of work $v(\cdot)$ is increasing, strictly convex and three times continuously differentiable. The utility function in (5) is separable between consumption and labour 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 parameterised tax schedules into the utility function, and optimising over labour income and savings, yields the following first-order conditions:

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

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

We denote marginal tax rates by a prime, so that $T'(z^n, \rho^T, \sigma^T) \equiv \partial T(z^n, \rho^T, \sigma^T)/\partial z^n$ and $\tau'(y^n, \rho^\tau, \sigma^\tau) \equiv \partial \tau(y^n, \rho^\tau, \sigma^\tau)/\partial y^n$. Other partial derivatives are denoted by a subscript. Thus, $u_1(\cdot)$ and $u_2(\cdot)$ are the marginal utility of first- and second-period consumption and $y_a(\cdot)$ denotes the marginal rate of return. Equation (6) shows that the marginal rate of substitution between first-period consumption and leisure must equal the marginal after-tax wage rate. Equation (7) shows that the marginal rate of substitution between first- and second-period consumption must equal the individual's discount factor. We define the inverse of the discount factor—or one plus the after-tax rate of return—as $R^n \equiv 1 + (1 - \tau')y_a$.[10]

[9] It is common to parameterise non-linear tax schedules to derive the comparative statics, see, e.g., Christiansen (1981), Immervoll *et al.* (2007), Jacquet *et al.* (2013) and Gerritsen (2016; 2024).

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

We impose a number of assumptions that help us derive the optimal non-linear tax schedules. First, we require that both tax schedules are three times continuously differentiable. This ensures that both the individual first-order conditions and the relevant behavioural elasticities are differentiable. Second, we assume that second-order conditions are satisfied and that (6) and (7) describe a unique and global maximum for utility. This guarantees that individual behaviour is differentiable and thus that marginal changes in taxes lead to marginal responses in earnings. These assumptions correspond to Assumption 2 in Jacquet and Lehmann (2021).

Third and final, we assume that the equilibrium values of both tax bases y^n and z^n are monotonically increasing in ability n. This implies a one-to-one mapping between labour and capital incomes, which allows us to derive the Pareto-efficient mix of taxes on capital and labour income. The one-to-one mapping ensures that the distributional impact of an increase in one tax instrument can always be replicated by an increase in the other. The Pareto-efficient tax mix then equates the marginal excess burdens associated with any given distributional impact for both taxes. Empirical evidence shows that labour income and capital income or wealth are indeed strongly positively correlated—lending some empirical support to the presumed one-to-one mapping between labour and capital income.[11] Nevertheless, this is a simplification of reality. We briefly return to this issue in the Conclusion.

2.2. *Behavioural Elasticities*

Behavioural elasticities of the tax bases play an important role in the optimal tax expressions that we derive below. To define these elasticities, we first write the tax bases as functions of the tax parameters. The first-order condition for savings in (7), together with the definitions of first- and second-period consumption in (1) and (2), implicitly determines equilibrium savings as a function of labour income, tax parameters and ability. This allows us to write equilibrium savings as $a^n = \tilde{a}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$, where the superscript c indicates conditionality on labour income z^n. Since capital income is a function of savings and ability, we can write equilibrium capital income as a function of the same arguments $y^n = \tilde{y}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n) = y(\tilde{a}^c, n)$. The two first-order conditions in (6) and (7), together with the definitions of first- and second-period consumption in (1) and (2), determine labour income as a function of tax parameters and ability. This allows us to write equilibrium labour income as $z^n = \tilde{z}(\rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$.[12]

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

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

The elasticity in (8) measures the percentage change in labour income if the net-of-tax rate $1 - T'$ is exogenously raised by one percent, while utility is kept constant. It captures the *total* impact on labour income, taking into account second-round effects due to the impact of a change in labour income on the marginal tax rate if the tax function is non-linear.[13] The term within brackets

[11] For example, Black *et al.* (2023) show that labour income is the most important determinant of wealth in Norway. Similarly, Kaymak *et al.* (2022) show that disparities in labour income are an important driver of wealth inequality in the United States.

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

[13] In terms of Jacquet and Lehmann (2021), e_z^n is a 'total elasticity' rather than a 'direct elasticity'. The elasticity measures the effect on labour income of a given change in the tax parameters σ^T and ρ^T rather than a given change in

gives the Slutsky decomposition of the compensated response in labour income to an increase in marginal taxes.[14]

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

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

The elasticity in (9) measures the percentage change in capital income if the after-tax rate of return $R^n = 1 + (1-\tau')y_a$ is exogenously raised by one percent, while utility is kept constant. It captures the total impact on capital income, while taking into account second-round effects due to the impact of a change in capital income on the marginal tax rate if the tax function is non-linear. Again, the term within brackets is the Slutsky decomposition of the compensated response of capital income to an increase in the marginal tax rate. Furthermore, $e_{y|z}^n$ is a *conditional* elasticity, since it measures the behavioural change in capital income while holding labour income constant.

In what follows, we only employ the elasticities of the tax bases for labour and capital income with respect to their 'own' marginal tax rates. Naturally, both labour and capital income are also affected by the 'other' marginal tax rates. The marginal tax rate on capital income affects labour income and vice versa. We do not need to explicitly define the associated cross-elasticities, because we can write all compensated cross-effects in terms of the compensated 'own' tax elasticities.[15]

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

$$\xi_z^n \equiv \frac{\partial \tilde{z}}{\partial n}\frac{n}{z^n}, \qquad \xi_{y|z}^n \equiv \frac{\partial \tilde{y}^c}{\partial n}\frac{n}{y^n}. \tag{10}$$

The first elasticity ξ_z^n measures the percentage change in labour income due to a one-percent increase in ability. The second elasticity $\xi_{y|z}^n$ measures the percentage change in capital income due to a one-percent increase in ability, *while holding labour income constant*. The elasticity $\xi_{y|z}^n$ captures the extent to which ability *directly* raises capital income, instead of indirectly through increased labour income. Thus, it gives a measure of the type dependence in capital returns. Alternatively, ξ_z^n measures the rents from ability that accrue in labour income, while $\xi_{y|z}^n$ measures the extent to which ability rents end up in capital income, after controlling for labour income. In the absence of type dependence, differences in capital income are perfectly explained by differences in labour income so that the conditional elasticity $\xi_{y|z}^n$ is zero in that case.

3. Two Micro-foundations of Return Heterogeneity

It is instructive to consider two plausible micro-foundations for capital income $y(a^n, n)$ that could generate heterogeneity in capital returns. These two different micro-foundations loosely correspond to what Gabaix *et al.* (2016) call type-dependent and scale-dependent returns.

the marginal tax rate $T'(z, \rho^T, \sigma^T)$. Total elasticities are also used by, e.g., Jacquet *et al.* (2013), Jacobs and Boadway (2014), Gerritsen (2016) and Scheuer and Werning (2017).

[14] We formally prove this in Online Appendix B.3.1.

[15] The reason for this is as follows. First, Slutsky symmetry allows us to write the compensated cross-effect of capital taxes on labour income in terms of the compensated cross-effect of labour taxes on capital income. See also Online Appendix B.3. Second, the marginal tax rate on labour income does not *directly* affect the incentives to save. Instead, the effect on capital income only runs through its impact on labour income and is, therefore, proportional to the elasticity of labour income with respect to the 'own' tax rate.

3.1. *Type-Dependent Returns: Entrepreneurial Investments*

We first consider type-dependent returns. In particular, we consider an economy in which individuals can invest in two different assets. They can invest in a closely held asset that is specific to their type and could be interpreted as entrepreneurial investment. And they can invest in an asset that is freely traded in capital markets. Individual n invests b^n in the closely held asset. This yields a total return that is a function of invested capital and ability: $\pi^n = \pi(b^n, n)$. The closely held asset exhibits decreasing returns to capital ($\pi_b > 0$ and $\pi_{bb} < 0$) and positive returns to ability ($\pi_n > 0$). The latter assumption reflects the idea that high ability helps to find and select successful business ventures. The remainder of individual savings, $a^n - b^n$, is invested in the freely traded asset, which yields a common, constant rate of return r.

Capital income is now given by:

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

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

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

Hence, the general formulation $y^n = y(a^n, n)$ can capture the special case of entrepreneurial investments. Under this micro-foundation, capital income is linear in savings and increasing in ability: $y_a = r$ and $y_n > 0$.[16] While the marginal rate of return (y_a) is identical for everyone, the average rate of return (y^n/a^n) typically varies across the population. As long as $y_n > 0$, the average rate of return on a given amount of savings is increasing in ability.

3.2. *Scale-Dependent Returns: Scale Economies in Wealth Accumulation*

The second micro-foundation of capital income $y(a^n, n)$ relies on scale economies in accumulating wealth. Scale economies may originate from fixed costs associated with realising higher rates of return. By fixed costs, we mean costs that do not vary with the amount of invested wealth. 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 invested wealth is sufficiently large. As a consequence, individuals with more wealth are likely to obtain higher rates of return.

We can capture scale effects in our model by assuming that individuals invest x^n of their savings to raise the returns on the remainder of their savings. These investments consist of search costs, fees and the costs of obtaining financial know-how. This leaves an amount $a^n - x^n$ to be saved at a rate of return $r(x^n) \geq 0$ with $r'(x^n) \geq 0$. We further assume that investments x^n raise the rate

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

of return at a decreasing rate, such that $r''(x^n) < 0$ and that the rate of return $r(x^n)$ is bounded from above. Although this latter assumption is much stricter than necessary, it is sufficient to guarantee that second-order conditions for the individual choice problem are satisfied. Moreover, it makes intuitive sense that rates of return cannot grow without bounds.

We assume that all investment costs are deductible from the capital tax base, but not separately observed by the government and thus not separately taxed.[17] Taxable capital income is then given by:

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

Individuals invest in financial services to maximise their capital income. Maximising y^n in (13) with respect to x^n yields $r'(x^n)(a^n - x^n) = 1 + r(x^n)$.[18] 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 its rate of return. Substituting this into the expression for capital income yields:

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

Hence, the general formulation $y^n = y(a^n, n)$ also captures scale economies in wealth accumulation. 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 $y_n = 0$.[19] Both the marginal rate of return (y_a) and the average rate of return (y^n/a^n) are (weakly) increasing with savings.[20]

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

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

[18] The second-order condition is given by $r''(x^n)(a^n - x^n) < 2r'(x^n)$ and it is satisfied by virtue of our assumptions on capital returns: $r'(x^n) > 0 > r''(x^n)$.

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

[20] The average rate of return is (weakly) increasing with savings if and only if the marginal rate of return (weakly) exceeds the average rate of return, such that $y_a \geq y^n/a^n$. This follows from the assumption that $y_{aa} \geq 0$.

4. Pareto-Efficient Taxation

4.1. *Social Welfare and Government Budget Constraints*

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

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

Social preferences for income redistribution are captured by concavity of either the social welfare function $\mathcal{W}$ or the utility function U^n. We use the social welfare function in (15) to characterise optimal tax schedules separately for labour and capital income. But it is irrelevant for our characterisation of the Pareto-efficient mix of the two taxes.

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

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

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

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

Fixing government assets through (16) and (17) is innocuous in the first micro-foundation with type-dependent returns. This is because government debt is neutral if the government can borrow and lend at the same marginal rate as all individuals.[21] We nevertheless fix government assets to ensure an interior solution for the optimal tax schedules in the case of scale-dependent returns. Without fixing government assets, the government may want to exploit scale effects to the maximum possible extent by taking over all investments in the entire economy. It could do so by letting the intercept of the labour (capital) income tax go to infinite (minus infinite). Fixing the government asset position should be seen as a simple short-cut to modelling the possible efficiency losses or political-economy distortions associated with managing extremely large public wealth funds.[22]

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

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

Equation (18) is written in units of first-period government revenue. It shows that the government discounts future tax revenue at a rate λ_1/λ_2.

[21] Ricardian equivalence applies even though taxes are distortionary, since the government has access to a non-distortionary marginal source of public finance in each period. Hence, the government does not need to introduce tax distortions to steer the inter-temporal allocation. See also Werning (2007) and Jacobs (2018).

[22] The assumption of fixed government assets only affects the optimal intercepts of the tax schedules. None of the derivations of the optimal marginal tax rates depend on this assumption.

4.2. *Instrument Set*

In deriving the optimal tax schedules on labour and capital income, we impose a number of restrictions on the instrument set of the government. First, we assume that the government cannot observe ability. This implies that the government cannot redistribute income by using non-distortionary individualised lump-sum taxes and transfers. Instead, the government has to rely on distortionary taxes on tax bases it *can* observe. In particular, the government can observe and therefore tax labour income and capital income at the individual level. Second, we follow most of the literature by assuming that capital income and labour income are taxed with separable schedules (e.g., Saez and Stantcheva, 2018). As a result, the marginal tax rate on one tax base does not depend on the size of the other tax base.[23]

Third, we assume that both income taxes are only levied in the same period in which income is earned. Thus, taxes on labour (capital) income are only levied in the first (second) period. This assumption is irrelevant when considering type-dependent returns, but crucial when considering scale-dependent returns. In the latter case, it is optimal to redistribute from rich to poor relatively late in the life cycle, allowing the rich to more fully exploit their scale advantages. This makes taxes on capital income attractive because capital income is earned and thus taxed relatively late in life. If, however, the government could tax first-period labour income with two different tax schedules in both the first *and* the second period, then it would no longer need taxes on capital income to shift the tax burden to the second period. A possible micro-foundation for the assumption that income is taxed when earned is that the government has limited ability to keep reliable tax records over long time intervals. It is also reflective of real-world income taxes that are typically only levied during (or soon after) the period in which the income is earned.

Finally, we rule out taxes on consumption and wealth as this would enable the government to tax away all excess returns with zero distortions.[24] In practice, 100% taxation of excess returns via wealth or consumption taxes will surely result in tax avoidance and evasion due to cross-border shopping, international mobility of capital and reduced entrepreneurial efforts. Hence, our model attempts to realistically capture the main policy trade-off for the optimal taxation of capital income with heterogeneous returns, while avoiding complexities with modelling cross-border shopping, capital mobility or entrepreneurial effort. In Section 4.7, we briefly consider how our results may change if the government were to tax wealth instead of capital income.

4.3. *Excess Burdens and Social Welfare Weights*

The optimal tax structure depends on the excess burdens and distributional benefits of taxes on labour and capital income. We define the marginal excess burden as the revenue loss caused by

[23] In the case of type-dependent returns, the assumption of separable tax schedules is not a binding constraint. Indeed, by following a first-order mechanism-design approach, we can show that the optimal incentive-compatible allocation derived from the direct mechanism can be implemented with separable tax schedules. These derivations are available upon request. Naturally, this presupposes that the first-order mechanism-design approach is valid and that taxpayers are not better off in bundles outside of the optimal allocation when the tax schedules are separable, and thus that preferences and technologies satisfy certain implementability conditions. These conditions are derived by Ferey *et al.* (2024); we simply assume they hold. In the case of scale-dependent returns, the assumption of separable tax schedules *is* a binding constraint. As we discuss below, the government could achieve higher levels of welfare were it able to set two joint tax schedules in both periods instead of two separable tax schedules.

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

a *compensated* increase in a marginal tax rate. The marginal excess burdens of taxes on labour and capital income for individual n are given by:

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

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

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

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

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

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

The social welfare weights in (21) and (22) consist of the welfare gains (in resource units) 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.

4.4. *Optimal Tax Schedules*

We solve for the optimal non-linear taxes on labour and capital income by using the Euler–Lagrange formalism, which is a mathematically rigorous version of the heuristic tax-perturbation approach pioneered by Saez (2001) and recently extended and amended by Golosov *et al.* (2014), Spiritus *et al.* (forthcoming) and Gerritsen (2024). In particular, the Euler–Lagrange formalism employs the calculus of variations to analyse the welfare effects of small perturbations in the tax schedules on labour and capital income. In the optimum, such tax perturbations should have no effect on social welfare.

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

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

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

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

PROOF. See Online Appendix B.1. □

The conditions in Lemma 1 are intuitively straightforward. Consider a small change to the marginal tax rate on labour income in a small interval around income z^n. This tax change distorts labour supply for all individuals with income around z^n. The excess burden associated with the distortion is given by the left-hand side of (23). The perturbation also raises tax revenue from individuals who earn more than z^n. The associated redistributional gains are given by the right-hand side of (23). Analogously, the left-hand side of (24) gives the marginal excess burden of raising the marginal tax rate on capital income around y^n, and the right-hand side gives the redistributional gains of doing so. In the optimum, the marginal excess burden of raising a marginal tax rate on either labour or capital income should thus be equal to its marginal redistributional gains.

4.5. *Optimal Taxation of Labour Income*

Lemma 1 expresses the optimal tax schedules in terms of marginal excess burdens and redistributional gains of taxation. To gain more insight into the shape of the optimal tax schedules, we write them in terms of wedges, elasticities, the income distribution and social welfare weights. The following proposition establishes the optimal tax wedge on labour 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 labour and capital income.

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

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

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

PROOF. See Online Appendix B.2. □

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

($s = 0$), reductions in labour income do not reduce future consumption, so that the standard Saez-formula results.

4.6. *The Pareto-Efficient Mix of Labour- and Capital-Income Taxes*

Proposition 1 shows how taxes on labour income should be set to optimally trade-off marginal distortions in labour supply against marginal redistributional benefits. However, given the one-to-one mapping between labour income and capital income, the same distributional benefits can also be obtained by taxing capital income. The Pareto-efficient tax structure requires that taxes on labour income and taxes on capital income yield the same marginal excess burdens for the same marginal income redistribution. Otherwise, a Pareto improvement could be obtained by redistributing a little more with one tax instrument and a little less with the other. This insight allows us to derive expressions for the Pareto-efficient tax mix. These expressions only depend on excess burdens (captured by tax wedges and elasticities), but not on the government's preference for income redistribution (captured by the social welfare weights).[25]

Below, we first consider the Pareto-efficient tax mix for type-dependent returns, then for scale-dependent returns. In Online Appendix A, we discuss the general formulation of the optimal tax on capital income that captures both micro-foundations as special cases.

4.6.1. *Type-dependent returns ($y_a = r,\ y_n \geq 0$)*

The following proposition presents an expression for the Pareto-efficient tax mix with type-dependent returns. This tax structure ensures that the marginal excess burdens of achieving a given redistribution are equalised across tax instruments.

PROPOSITION 2. *If capital returns are type dependent ($y_a = r$ and $y_n \geq 0$ for all individuals n), then any Pareto-efficient tax mix requires capital and labour income to be taxed such that:*

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

for every level of capital income y and corresponding labour income z. The inequality is strict only if $\xi_{y|z} > 0$, which holds if and only if $y_n(a, n) > 0$.

PROOF. See Online Appendix B.3. □

The first striking implication of Proposition 2 is that Pareto-efficient marginal taxes on capital income are strictly positive if returns increase in ability, i.e., if $y_n(a, n) > 0$. With type-dependent returns, capital income reflects both previously earned labour income *and* ability rents. A tax on capital income ensures that these rents, which escape the tax on labour income, are taxed. Since taxing rents is non-distortionary, the existence of ability rents in capital income reduces the efficiency costs associated with taxing capital income. As a result, for the same distributional effect, a tax on capital income distorts labour supply less than a tax on labour income. While taxing capital income also distorts savings decisions, these distortions are of second order if taxes on capital income are zero. Strictly positive tax rates on capital income are therefore part of any Pareto-efficient tax mix.

[25] See Koehne and Sachs (2022) for a similar approach in the context of Pareto-efficient tax deductions.

Taxing capital income, rather than labour income, thus *reduces* distortions of labour supply, while *raising* distortions of savings. The Pareto-efficient tax mix captures this trade-off, as illustrated by (26). The left-hand side gives the marginal excess burden on savings due to the tax on capital income. It equals the tax wedge on savings ($y_a\tau'(y)/(1+y_a)$), multiplied by the elasticity of capital income with respect to the after-tax interest rate ($e_{y|z}$). The right-hand side gives the marginal *reduction* in the excess burden on labour supply. The two terms in brackets represent the total tax wedge on labour income. Multiplied by the compensated elasticity of labour income (e_z), it gives the marginal excess burden of a tax on labour income. The ratio of ability elasticities ($\xi_{y|z}/\xi_z$) represents the importance of ability rents in capital income relative to labour income. Taxes on capital income are *less* distortionary if ability rents in capital income are relatively *more* important. Hence, the ratio of ability elasticities reflect the degree to which taxes on capital income distort labour supply less than a tax on labour income.

Equation (26) gives a condition for the Pareto efficiency of the tax system. As a result, it does not in any way depend on the shape of the social welfare function. Instead, we find that the Pareto-efficient tax wedge on savings depends on a limited number of empirically measurable sufficient statistics: elasticities and tax wedges. First, it is decreasing in the compensated elasticity of capital income ($e_{y|z}$). This elasticity raises the costs of savings distortions and thus the relative costs of employing capital-income taxes. Second, it is increasing in the compensated elasticity of labour income (e_z) and the tax wedge on labour income. Both raise labour-supply distortions and thus the relative benefits of capital-income taxes. Finally, it is increasing in the ratio of ability elasticities ($\xi_{y|z}/\xi_z$).[26] This ratio captures the importance of ability rents in capital income, and represents the degree to which a tax on capital income distorts labour supply *less* than a tax on labour income. This term is crucial for our results. It shows that more type-dependent return heterogeneity—and thus more ability rents in capital income—is associated with higher optimal tax rates on capital income.

Proposition 2 nests the zero-tax result of Atkinson and Stiglitz (1976) as a special case. Only in the absence of return heterogeneity, such that $y_n(a,n)=\xi_{y|z}=0$, do we find that Pareto-efficient taxes on capital income are zero. In that case, taxes on capital income and taxes on labour income yield the same marginal excess burden on labour supply for the same redistribution. A tax on capital income then only generates savings distortions without having any benefits over a tax on labour income.

Proposition 2 is related to a number of other studies on the optimal taxation of commodities or capital income. Our result that capital income should be taxed if $\xi_{y|z}>0$ is in line with Mirrlees (1976), who argues that optimal commodity taxation depends on 'the way in which demands change *for given income and labour supply* when n changes'. This finding is echoed in subsequent studies. Christiansen (1984), Golosov *et al.* (2013), Jacobs and Boadway (2014) and Hellwig and Werquin (2024) show that non-separabilities in the utility function between consumption and leisure may generate reasons to differentiate commodity taxes or to tax capital income. Mirrlees (1976), Saez (2002) and Diamond and Spinnewijn (2011) obtain the same

[26] One way to estimate the conditional ability elasticity $\xi_{y|z}$ in (10) is by regressing capital income on wage rates as a proxy for ability, while controlling for labour income. Doing so, Gordon and Kopczuk (2014) find a positive correlation, which implies that $\xi_{y|z}/\xi_z>0$.

implication if preferences are heterogeneous across types.[27,28] Contrary to these earlier studies, we show that the argument for positive taxes on capital income does not rely on preferences being heterogeneous across agents or non-separable between consumption and leisure. Instead, the case for positive capital-income taxes can be driven by empirically plausible return heterogeneity.[29]

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

With *scale-dependent* returns, anyone can obtain higher rates of return simply by saving more—regardless of their ability. As a result, there are no ability rents in capital income. This means that labour-income taxes and capital-income taxes yield the same labour-supply distortions for the same redistribution—in line with Atkinson and Stiglitz (1976). Nevertheless, we show that a tax on capital income is still desirable, because it helps to improve the allocative efficiency of savings. The following proposition presents an expression for the Pareto-efficient tax mix with scale-dependent returns. It ensures that the marginal excess burdens of achieving a given redistribution are equalised across tax instruments.

PROPOSITION 3. *If capital returns are scale-dependent ($y_{aa} \geq 0$ and $y_n = 0$ for all individuals), then any Pareto-efficient tax mix requires capital income to be taxed such that:*

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

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

PROOF. See Online Appendix B.3. □

To understand Proposition 3, recall that wealthy individuals obtain higher marginal rates of return on capital than poor individuals. It would, therefore, be efficient if the rich were to save on behalf of the poor. But an implicit market failure prevents these transactions from taking place. As a result, savings are inefficiently allocated across the rich and the poor. The tax system may alleviate the market failure by reallocating savings from the poor to the rich.

Consider raising the marginal tax rate on labour income around income level z, to finance a reduction of the intercept $T(0)$. This raises the *first-period* tax burden for the relatively rich, with income above z, and reduces the first-period tax burden for the relatively poor, with income below z. Thus, while the reform only raises *marginal* tax rates for people with income z, it adjusts

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

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

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

[28] As mentioned in Section 1, our model with type-dependent returns is mathematically isomorphic to a model of heterogeneous preferences to save. Also see Online Appendix D.

[29] In our case, budget constraints rather than preferences vary with n for given labour income. In this respect, our findings are also related to Cremer *et al.* (2001), who find that taxes on capital income are desirable if endowments—which are part of the budget constraint—are increasing with ability.

average tax rates for everyone. The changes in tax burdens generate income effects on savings. The rich will reduce savings to smooth out their increase in first-period taxes. And the poor will increase savings to smooth out their reduction in first-period taxes. Taxes on labour income thus *exacerbate* the market failure by reallocating savings from the rich to the poor.

Now consider raising the marginal tax rate on capital income around y, to finance a reduction of the intercept $\tau(0)$. This raises the *second-period* tax burden for the relatively rich and reduces the *second-period* tax burden for the relatively poor. This again yields income effects on savings. As the change in tax burdens takes place in the second period, the rich smooth it out by saving *more*, while the poor will save *less*. Thus, in contrast to taxes on labour income, taxes on capital income *alleviate* the market failure by reallocating savings from the poor to the rich. The timing of the two tax instruments is crucial: taxes on capital income are levied later in life than taxes on labour income. By raising revenue later in life, taxation of capital income allows the rich to better exploit their scale effects in wealth accumulation.

While marginal taxes on capital income generate beneficial income effects, they also yield substitution effects, which cause distortions in savings. In particular, a marginal tax rate around y distorts savings behaviour for people who earn capital income around y. The Pareto-efficient tax mix balances the efficiency gains of alleviating the market failure with the efficiency costs of savings distortions. This is shown by (27).

The left-hand side is the tax wedge on savings for individuals with capital income y. Multiplied by $e_{y|z}yg(y)$ (the right-hand side denominator), it gives the marginal excess burden on savings of raising marginal taxes on capital income around y. The remaining right-hand side terms give the efficiency gains of alleviating the market failure. Higher marginal taxes around y take away $1 - G(y)$ resources from individuals with higher capital income. This induces the rich to smooth consumption by saving more at a rate of return y_a^+. The tax revenue can be redistributed to all individuals in the population by adjusting the intercept of the tax schedule $\tau(0)$. This induces everyone to smooth consumption by saving less at an average rate of return $\bar{y}_a$. By adding both effects, the efficiency gains of reallocating savings are thus given by the difference in rates of return $(\bar{y}_a^+ - \bar{y}_a)$, discounted at the government's discount rate $(\bar{y}_a)$.[30]

The expression in (27) thus gives the marginal efficiency costs and benefits of shifting the tax burden from labour to capital income in a distributionally neutral way. Such a shift increases savings distortions (left-hand side) and reduces the efficiency costs associated with the market failure (right-hand side). In contrast to Proposition 2, the tax wedge on labour income is absent from (27). This is because there are no type-dependent returns. As a result, a distributionally neutral shift of taxes from labour to capital income is neutral with respect to labour-supply distortions.

Equation (27) is a condition for Pareto efficiency and therefore does not depend on the shape of the social welfare function. The Pareto-efficient marginal tax rate on capital income only depends on a number of sufficient statistics with clear empirical counterparts. First, the optimal tax rate on capital income is decreasing in the compensated elasticity of capital income with respect to the after-tax interest rate $(e_{y|z})$. Second, it is decreasing in the concentration of capital income around y $(yg(y))$. Both terms raise the distortions of the marginal tax rate on capital income y. Third, the optimal tax on capital income increases in the share of individuals with income above y $(1 - G(y))$. A higher share implies that a marginal tax rate raises more revenue, resulting in larger income effects on savings, and thus a bigger improvement in the allocation of savings.

[30] Online Appendix B.3 formally shows that the government's discount rate indeed equals the average discount rate in the population.

Fourth, and most importantly, marginal taxes on capital income are increasing in the amount of return heterogeneity (as measured by $(\bar{y}_a^+ - \bar{y}_a)/(1+\bar{y}_a)$). Larger return heterogeneity implies that there are larger efficiency gains from making the rich save on behalf of the poor. It therefore raises the allocative efficiency gains of taxing capital income.[31]

All results on optimal second-best taxation are fundamentally driven by constraints on the government's instrument set that preclude a first-best outcome. Propositions 1 and 2 followed from the government's inability to directly tax ability. Proposition 3 follows from the difference in timing between labour-income taxes and capital-income taxes. With scale-dependent returns, it is efficient to tax the rich relatively late in life. If the government is able to perfectly and indefinitely keep tax records, it can tax first-period labour income with different schedules in both periods. It could then alleviate the market failure without imposing distortions on savings. However, if the government cannot perfectly and indefinitely keep tax records, it is restricted to tax income when it is earned. It is then optimal to set positive taxes on capital income, which is typically earned later in life than labour income.

To the best of our knowledge, this justification for positive taxes on capital income is entirely novel. Gahvari and Micheletto (2016) find no role for taxes on capital income if return heterogeneity stems from economies of scale. This follows directly from their assumption that both taxes on labour income and capital income are levied in the same period. Proposition 3 shows that their result breaks down if taxes on capital income are levied later in life than taxes on labour income.[32]

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

COROLLARY 1. *Assume that (i) returns are scale dependent ($y_{aa} \geq 0$ and $y_n = 0$ for all individuals), (ii) the right tail of the distribution of capital income follows a Pareto distribution and (iii) the conditional elasticity of capital income and the marginal rate of return on savings converge to the constants $\hat{e}_{y|t}$ and $\hat{y}_a$ for high levels of income. Then, the Pareto-efficient tax rate on capital income $\hat{y}$ at the top of the income distribution is constant and given by:*

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

where $p = \hat{y}g(\hat{y})/(1 - G(\hat{y}))$ is the Pareto parameter of the right tail of the distribution of capital income.

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

[31] Equation (27) resembles the ABC-formula of optimal non-linear taxes on *labour* income (see Proposition 1, as well as Diamond, 1998; Saez, 2001). The main contrast is that the difference in average welfare weights $(1 - \bar{\alpha}_1^+)$, which measures the social benefits of income redistribution, is replaced by the difference in marginal rates of return $((\bar{y}_a^+ - \bar{y}_a)/(1+\bar{y}_a))$, which measures the social benefits of alleviating the capital market failure. The economic intuition for the formula is otherwise similar.

[32] Erosa and Gervais (2002) and Conesa *et al.* (2009) propose positive taxes on capital income if taxes on labour income cannot be conditioned on age. However, their reasoning is very different from ours and relies on the relative complementarity of consumption at different dates with leisure (cf. Corlett and Hague, 1953; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018). The complementarity of leisure with consumption at different dates does not play a role in our model, since we assumed (weakly) separable preferences.

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

Equation (28) allows us to make a back-of-the-envelope calculation for the optimal top tax rate on capital income. To quantify the optimality condition, we rely on our simulations as described in the next section. There, we obtain $\bar{y}_a/\hat{y}_a = 0.79$, $\hat{e}_{y|z} = 0.41$ and $p = 2.73$ for the top decile.[33] The return heterogeneity follows from Fagereng *et al.* (2020), and the elasticity of capital income lies within a range of empirically plausible elasticities.[34] Imposing these values yields a substantial Pareto-efficient top tax rate of about 19%.

4.7. *Extension: The Pareto-Efficient Wealth Tax*

So far we have concentrated our discussion on the Pareto-efficient taxation of capital income. Recently, the merits and demerits of a tax on wealth have been discussed in several publications (e.g., Saez and Zucman, 2019; Scheuer and Slemrod, 2021; Guvenen *et al.*, 2023). It is therefore useful to analyse how our results change if the government were to impose a tax on wealth instead of on capital income. We assume that a wealth tax is levied on the sum of savings and capital income ($w^n \equiv a^n + y^n$). In case of type-dependent returns, ability rents are only contained in capital income and not in savings. A wealth tax is therefore less well targeted to tax ability rents than a tax on capital income. As a result, for a given tax revenue, the wealth tax is more distortionary than a tax on capital income.

Nevertheless, in Online Appendix E, we formally show that our qualitative results on Pareto-efficient taxes on capital income carry over to a setting with wealth taxes if consumption goods in both periods are normal. That is, for either type- or scale-dependent returns, positive taxes on wealth are part of the Pareto-efficient tax mix for the same reason why taxes on capital income are. A wealth tax is desirable because its base contains ability rents (in case of type-dependent returns) or because it effectively makes the rich save on behalf of the poor (in case of scale-dependent returns). The expressions for the Pareto-efficient mix of taxes on labour income and wealth are nearly identical to those in Propositions 2 and 3. The main difference is that elasticities and tax wedges now refer to wealth rather than capital income.

5. Numerical Simulation

In this section, we provide numerical simulations for the Pareto-efficient tax mix for the United States. In particular, we start from the actual US tax system and adjust it in Pareto-improving directions until no further Pareto improvements can be made. We do this separately for an economy with either type- or scale-dependent returns to capital. The Pareto-efficient US tax structure yields the same distribution of utilities as the *actual* US tax schedules on labour

[33] The actual Pareto parameter may well be smaller than in our simulations. For example, Vermeulen (2018) finds lower values for the Pareto parameter of wealth itself. The reason is that our calibration only takes heterogeneity in labour earnings and capital returns into account, whereas there are other reasons for concentration of capital income at the top in the data, such as differences in risk-taking, initial wealth, or preferences to save.

[34] See Seim (2017), Zoutman (2018) and Jakobsen *et al.* (2020) for a wide range of estimates.

and capital income, and only generates an increase in government revenue.[35] This increase in government revenue is the social welfare gain of moving from the current to the Pareto-efficient tax mix.

5.1. *Calibration of the Model*

The calibration of our model consists of a number of elements. First, we calibrate our two-period model to yearly data by assuming that each of the two periods consists of thirty-two identical years. Individuals make one labour-supply decision and stick to this decision for the first thirty-two years of life—yielding the same z^n in each year. Individuals also make one decision on how much of their after-tax labour income to save for retirement and stick to this savings decision for the first thirty-two years of life. Capital income is only taxed on realisation, i.e., when assets are sold in the final thirty-two years of life. During that period, individuals are retired and simply consume their savings and their after-tax capital income. Because the two periods are of equal length and all years within a period are identical, it is *as if* an individual consumes in year k of the second period the savings and their after-tax returns from year k of the first period. This allows us to retain the two-period structure of our theoretical model while calibrating it on empirically observable annual data.

Second, given a linear approximation of the US tax system, we calibrate the ability distribution in our model such that the resulting distribution of labour income approximates its empirical counterpart in the United States. We assume that ability n follows a log-normal distribution up to a certain level of ability n^*, after which it follows a Pareto distribution. Thus, $\ln n \sim N(\mu, \text{sd})$ for $n \in (0, n^*)$, where mean μ and standard deviation sd of log ability are chosen such that the mean and the median of labour income match their 2018 US values of \$54,906 and \$40,453 (US Census Bureau, 2019). We append the ability distribution with a Pareto tail for ability levels above n^*. The Pareto parameter of the right tail of the ability distribution is set to $p = 2.5$.[36] We choose n^* and the scale parameter of the Pareto tail such that the probability density function and its first derivative are continuous.

Third, we calibrate the capital-income function such that the resulting return heterogeneity matches Norwegian estimates from Fagereng *et al.* (2020).[37] In particular, we assign levels of capital income (y^n), returns to ability ($y_n(a^n, n)$) and returns to wealth ($y_a(a^n, n)$) to each level of ability n. We assume that capital income y^n varies across the wealth distribution according to the following functional form:

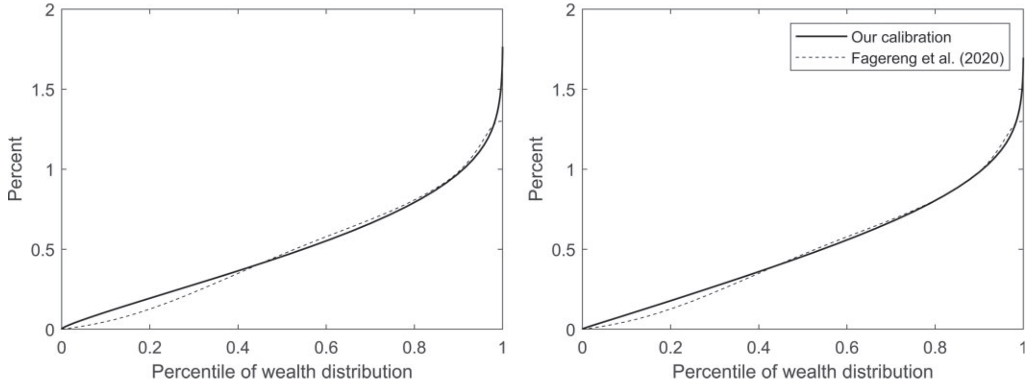
$$y^n = ra^n + \delta \left(\frac{a^n}{a^n - 1} \right)^{\rho} a^n, \quad \delta > 0, \quad \rho > 0. \tag{29}$$

This implies that average rates of return vary from r to $r + \delta$ as wealth increases from 0 to infinite. We choose this functional form because it provides for a good fit with the empirical estimates of return heterogeneity from Fagereng *et al.* (2020). As y^n in (29) corresponds to thirty-two-year

[35] The second-best Pareto frontier contains the full set of Pareto-efficient tax schedules. Each point on this Pareto frontier is associated with a different distribution of total tax burdens, which is determined by the Pareto-efficient mix of taxes on labour income and capital income.

[36] We use a somewhat higher value for the Pareto parameter of labour income than conventional estimates of around 2, since the latter apply to total income (Atkinson *et al.*, 2011). Since labour income is more equally distributed than capital income, the Pareto parameter for labour income is likely a bit higher.

[37] This reliance on a combination of US and Norwegian data is in line with other recent studies that numerically simulate heterogeneous returns on capital (e.g., Guvenen *et al.*, 2023).



(a) Skill-dependent returns

(b) Scale-dependent returns

Fig. 1. *Annualised Average Rate of Return, Relative to Poorest Individual, for Type- and Scale-Dependent Returns.*

compounded capital income, the implied annualised rate of return is given by:

$$\bar{r}^{annual} \equiv \left(1+\frac{y^n}{a^n}\right)^{1/32} - 1 = \left(1+r+\delta\left(\frac{a^n}{a^n-1}\right)^{\rho}\right)^{1/32} - 1. \tag{30}$$

We match r to an annualised rate of return of 3%, which is close to the average return on a thirty-year Treasury bill in the last ten years (Federal Reserve Bank of St. Louis, 2020). Thus, we set $r = 1.03^{32} - 1 = 1.58$. We estimate δ and ρ in (30) by minimising the sum of squared differences between $\bar{r}^{annual}$ and empirical risk-adjusted rates of return for each level of ability. We obtain the latter from Fagereng *et al.* (2020), who find that risk-adjusted rates of return increase with about 1.3 percentage points from the poorest to the richest percentile.[38]

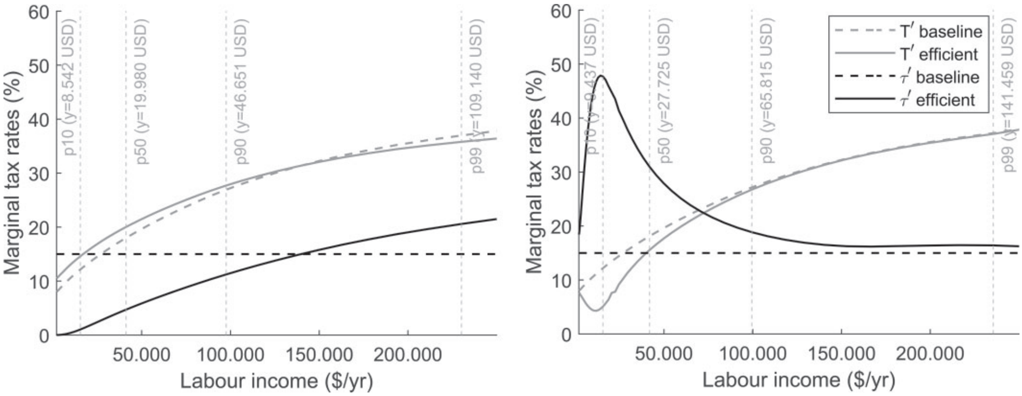
Levels of capital income y^n then follow from (29). For type-dependent returns, we set $y_a = r$ and we equate y_n to the numerical derivative of excess returns, $y^n - ra^n$, with respect to ability. In case of scale-dependent returns, we set $y_n = 0$ and determine y_a by differentiating (29). Figure 1 shows that the resulting return heterogeneity closely matches that of Fagereng *et al.* (2020). Because it is unclear to what extent the findings of Fagereng *et al.* (2020) can be extrapolated to the United States, we also consider more conservative cases with less return heterogeneity in our robustness analyses.

Fourth, we calibrate a utility function with a constant elasticity of inter-temporal substitution and a constant Frisch elasticity of labour supply:

$$U^n = \frac{c_1^{1-1/\sigma}}{1-1/\sigma} + \beta\frac{c_2^{1-1/\sigma}}{1-1/\sigma} - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}, \quad \beta, \varepsilon, \sigma > 0, \tag{31}$$

where σ is the elasticity of inter-temporal substitution, ε is the Frisch elasticity of labour supply and β captures the time preference of individuals. We set $\beta = 1/(1+r)$. In our baseline simulations, we adopt a Frisch elasticity of labour supply equal to $\varepsilon = 0.22$. This is similar to Saez (2001) and in line with empirical estimates discussed in Blundell and MaCurdy (1999) and Meghir and Phillips (2010). The elasticity of inter-temporal substitution (EIS) is set to $\sigma = 0.5$.

[38] See Fagereng *et al.* (2020, Fig. 3). We match estimates for the 10th, 50th, 90th and 99th percentiles of the wealth distribution and use a cubic interpolation for the remainder of the wealth distribution.



(a) Type-dependent returns
(b) Scale-dependent returns

Fig. 2. *Pareto-Efficient Versus Baseline Taxes for Type- and Scale-Dependent Returns.*

There is a wide range of empirical estimates for the EIS, but the vast majority of the estimates are well below 1. See Attanasio and Weber (2010) and Best *et al.* (2020) for reviews of the literature, and Havránek (2015) for a meta study. We consider the sensitivity of our results to alternative values of the Frisch elasticity and the EIS.

Finally, we derive baseline taxes on labour and capital income from the actual US tax system. We retrieve actual marginal tax rates on labour income from the online TAXSIM tax calculator of the National Bureau of Economic Research (Feenberg and Coutts, 1993). We do so for taxpayers who live in California, are single and have no dependants. Our baseline $T'(z)$ matches the smoothed schedule of actual marginal tax rates, featuring an average income-weighted tax rate of 23% and an average tax rate in the top decile of 33%. The baseline marginal tax rate on capital income $\tau'(y)$ is set to 0.15, the middle bracket tax rate of the tax on long-term capital gains in the United States.[39] This yields the baseline tax schedules as shown in Figure 2. The intercept of the labour-income tax schedule is adjusted such that the government raises 10% of GDP in exogenous revenue. The intercept of the capital-income tax schedule is normalised to zero.

In Online Appendix C, we provide a description of the algorithm by which we calibrate our model.

5.2. *Simulation Method*

Imposing the baseline tax schedules on our model yields a distribution of utilities across all individuals. In our simulations, we vary the tax schedules to obtain the tax mix that delivers this baseline distribution of utilities in the most efficient way. We solely focus on adjustments of marginal tax rates by keeping intercepts fixed.[40] The Pareto-efficient tax mix is computed by solving for the marginal tax rates on capital and labour income for each ability level such

[39] Our theoretical model does not distinguish between capital gains and other forms of capital income that are taxed as ordinary income in the United States (e.g., interest income and unqualified dividend income). We therefore also ran our simulations with a baseline tax rate on capital income of 0.23. This naturally causes both Pareto-efficient tax schedules to shift upwards, but does not affect any of our qualitative results.

[40] We do this because, in the case of scale-dependent returns, optimisation over intercepts may not yield an interior solution. Also see the discussion in Section 4.1.

that (*i*) the Pareto efficiency condition in (26) (for type-dependent returns) or (27) (for scale-dependent returns) is satisfied and (*ii*) each individual obtains the same utility (31) as in the baseline calibration. The resulting tax schedules deliver the baseline distribution of utilities with the lowest possible excess burden. The efficiency gain of implementing the Pareto-efficient tax mix is measured by the increase in discounted government revenue.

5.3. *Pareto-Efficient Taxes on Capital Income*

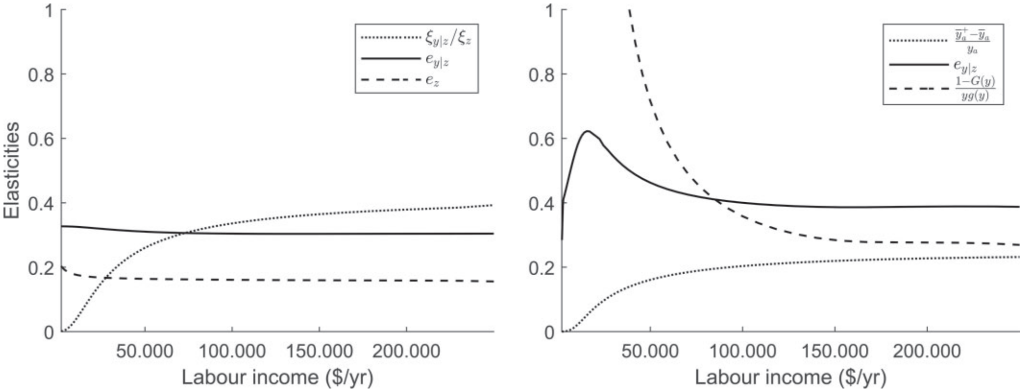
Figure 2 plots the Pareto-efficient tax schedules on labour and capital income, departing from the current US tax system. Panel (a) does this for type-dependent returns and Panel (b) for scale-dependent returns. In each panel, the solid (dashed) lines correspond to the Pareto-efficient (actual) tax schedules. The black (grey) lines correspond to marginal taxes on capital income (labour income). All tax schedules are plotted against labour income on the horizontal axis, so that marginal tax rates on both labour and capital income can be inferred for each individual and income percentiles can be indicated with the same dashed vertical lines. For the 10th, 50th, 90th and 99th percentiles, we give the associated level of capital income in parentheses.[41] Relative to the baseline, Pareto-efficient taxes on labour and capital income move in opposite directions. This is true by construction because utility remains constant. Since the change in the labour tax schedule is simply the mirror image of the change in the capital tax schedule, we focus our discussion on the Pareto-efficient tax schedule on capital income.

Figure 2 illustrates the two main insights from our numerical simulations. The first is that Pareto-efficient marginal tax rates on capital income are positive and economically significant for a majority of taxpayers. This is true for both type- and scale-dependent returns. The income-weighted average of the Pareto-efficient marginal tax on capital income is 9.6% with type-dependent returns and 25.1% with scale-dependent returns. The quantitative importance of the Pareto-efficient tax on capital income is also apparent at the top of the income distribution. The income-weighted average Pareto-efficient tax rate on capital income for the top decile is equal to 17.7% with type-dependent returns and 16.8% with scale-dependent returns. These are economically significant tax rates, particularly considering the fact that our model disregards all other relevant reasons to tax capital, as surveyed in Section 1. Accounting for additional reasons to tax capital income may imply even higher Pareto-efficient tax rates. Nevertheless, return heterogeneity alone is sufficient to generate Pareto-efficient taxes on capital income that are of the same order of magnitude as, or even considerably higher than, current US taxes on capital gains.

The second main insight is that the shape of the Pareto-efficient tax schedule on capital income strongly depends on whether returns are type- or scale-dependent. With type-dependent returns, Pareto-efficient marginal tax rates on capital income are monotonically increasing in income. This is primarily driven by two factors in the expression for Pareto-efficient taxes in (26).

First, the Pareto-efficient marginal tax rate on capital income is increasing in the ratio of ability elasticities, $\xi_{y|z}/\xi_z$, which captures the importance of ability rents in capital income. Indeed, our calibration of return heterogeneity implies that this ratio increases monotonically over the income distribution, as shown in Figure 3(a). Second, the optimal marginal tax rate on

[41] These levels of capital income cannot readily be compared to conventional statistics on the distribution of capital income. This is because we define capital income as the 32-year compounded returns on one year's worth of savings, while available statistics normally define it as the yearly returns on accumulated savings from all previous years. See Online Appendix C for more details on our calibration.



(a) Type-dependent returns (b) Scale-dependent returns

Fig. 3. *Key Drivers of the Pareto-Efficient Tax on Capital Income for Type- and Scale-Dependent Returns.*

capital income is increasing in the tax wedge on labour income $T'(z)/(1 - T'(z))$. This tax wedge is also monotonically increasing with income, as seen in Figure 2(a). Hence, the Pareto-efficient tax mix features increasing marginal tax rates on capital income. In fact, Pareto-efficient taxes on capital income exceed the actual US tax on capital gains only for the richest people.

This is entirely different if return heterogeneity originates from scale-dependent returns. In that case, Pareto-efficient taxes on capital income exceed the actual US tax on capital gains for virtually everyone, *particularly the less affluent ones*. Figure 2(b) shows that the Pareto-efficient tax schedule on capital income is first steeply increasing and then decreasing, i.e., it features a hump shape. This shape is also driven by two factors in the expression for Pareto-efficient taxes, see (27).

First, the Pareto-efficient tax on capital income is increasing in the relative difference in marginal rates of return between the rich and the average person, $(\bar{y}_a^+ - \bar{y}_a)/y_a$. This term is zero at the bottom, and then monotonically increases with income until it flattens out again, as seen in Figure 3(b). Second, the Pareto-efficient tax rate is increasing in the inverse relative hazard rate of the capital income distribution, $(1 - G(y))/(yg(y))$. Figure 3(b) shows that the inverse relative hazard rate is mostly declining with income before flattening out. Together these two terms explain the hump-shaped Pareto-efficient tax schedule on capital income. The effects of return heterogeneity and the relative hazard rate on the shape of the tax schedule overturn the impact of the elasticity of capital income, $e_{y|z}$. The latter is hump-shaped in income and would thus ceteris paribus lead to a U-shaped tax schedule, see Figure 3(b). The non-linear shape of the elasticity is most likely due to the non-linearity of the tax on capital income.

The difference between type- and scale-dependent returns is even more striking if we compare Pareto-efficient taxes on capital income with those on labour income. In case of type-dependent returns, Pareto-efficient taxes on capital income are smaller than those on labour income. This is in line with what we typically observe in most countries. However, in case of scale-dependent returns, Pareto-efficient taxes on capital income exceed those on labour income for a large majority of the population. These differences underscore the importance of further empirical research into the exact mechanisms behind return heterogeneity.

Our analysis suggests that it is possible to achieve a Pareto improvement by shifting the tax burden between labour and capital income. However, the welfare gains of moving from the current US system of taxes on labour and capital income to the Pareto-efficient tax system are only modest. Discounted government revenue can be increased by 0.013% of baseline GDP for type-dependent returns and by 0.047% of baseline GDP for scale-dependent returns. This may be because the Pareto-efficient tax structure does not deviate that much from actual taxes—which is especially visible if we compare actual and Pareto-efficient taxes on labour income in Figure 2.

5.4. *Robustness Analyses*

To assess the robustness of our results, we perform a number of simulations with alternative assumptions on the amount of return heterogeneity, the Frisch elasticity of labour supply and the EIS. For each robustness check, we re-calibrate the model to match our empirical targets for the labour income distribution and the excess returns on capital. We present our findings for the Pareto-efficient non-linear capital-income tax in Figure 4.[42] The left (right) panels refer to simulations with type-dependent (scale-dependent) returns. The first row shows results if we consider either 30% or 60% less return heterogeneity than in our baseline. The second row varies the Frisch elasticity. The third row varies the EIS.

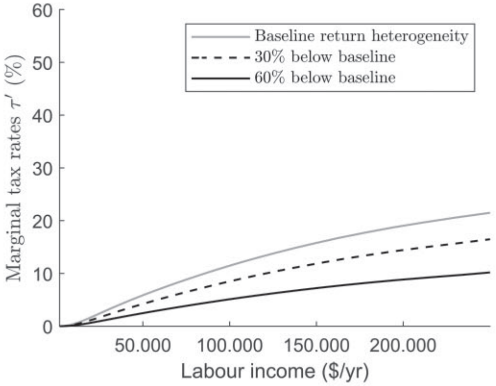
Our results are intuitive. Lower degrees of return heterogeneity lead to lower Pareto-efficient tax rates on capital income (Figure 4(a) and (b)). After all, in our model, return heterogeneity is the only reason to have taxes on capital income. A higher value of the Frisch elasticity raises Pareto-efficient tax rates on capital income with type-dependent returns (Figure 4(c)). With type-dependent returns, taxes on labour income distort labour supply more than taxes on capital income for the same redistribution of income. This relative disadvantage of taxes on labour income is exacerbated by a higher Frisch elasticity. The same is not true for scale-dependent returns, in which case taxes on labour income and capital income generate the same distortions in labour supply. A change in the Frisch elasticity then alters Pareto-efficient taxes on capital income mainly through the re-calibration of the model, which changes the elasticity of capital income and the relative hazard rate of the income distribution. As a result, the Frisch elasticity has an ambiguous effect on capital-income taxes with scale-dependent returns (Figure 4(d)). Finally, a higher value of the EIS reduces Pareto-efficient taxes on capital income (Figure 4(e) and (f)). This is because taxes on capital income are more distortionary with a higher EIS.[43]

6. Conclusion

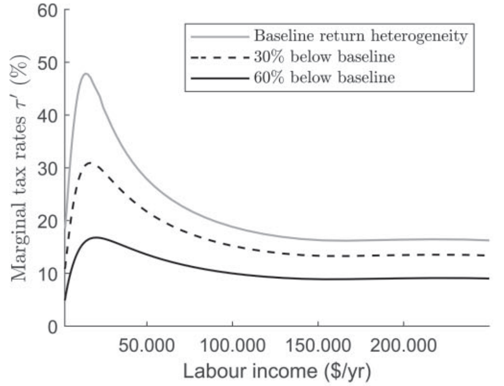
This paper analyses how the burden of taxation should be distributed between labour and capital income if individuals differ in their rates of return on capital. We show that any Pareto-efficient tax mix features positive taxes on capital income that are increasing in the degree of return

[42] Results for the Pareto-efficient labour-income tax are available on request from the authors.

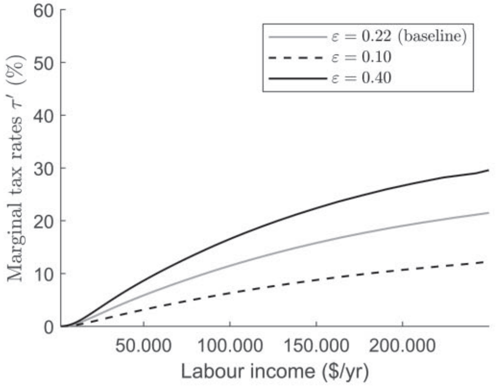
[43] An earlier version of this paper, available from the authors' websites, also provides simulations of the Pareto-efficient tax mix that maximises a standard social welfare function. It does so for type-dependent returns and also explores optimal restricted tax systems: one in which taxes are solely levied on labour income and one in which taxes are solely levied on comprehensive income, defined as the sum of labour and capital income. These simulations confirm that optimal marginal tax rates on capital income are quantitatively substantial. Moreover, welfare losses of not taxing capital income at all are of the same order of magnitude as taxing comprehensive income. However, because the simulations impose a specific social welfare function, the resulting tax schedules do not constitute a Pareto improvement over actual tax policy.



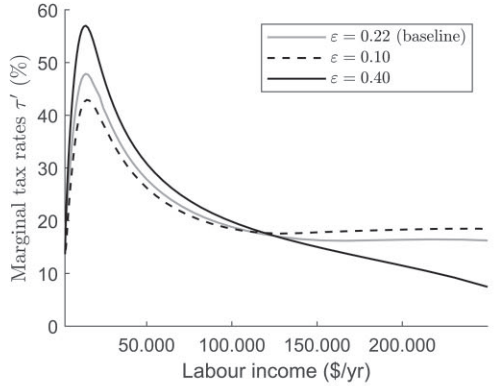
(a) Type-dependent returns; varying the degree of return heterogeneity



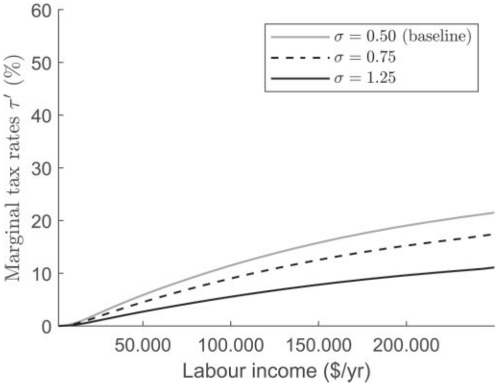
(b) Scale-dependent returns; varying the degree of return heterogeneity



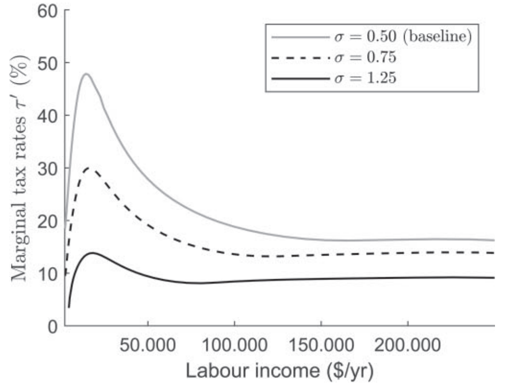
(c) Type-dependent returns; varying the Frisch elasticity



(d) Scale-dependent returns; varying the Frisch elasticity



(e) Type-dependent returns, varying the EIS



(f) Scale-dependent returns, varying the EIS

Fig. 4. *Robustness Analyses with Respect to Return Heterogeneity (a,b), the Frisch Elasticity (c,d) and the Elasticity of Inter-Temporal Substitution (e,f), for Type- (a,c,e) and Scale-Dependent Returns (b,d,f).*

heterogeneity, regardless of whether returns are type dependent or scale dependent. An empirically plausible numerical simulation of our model results in a Pareto-efficient tax mix with substantial taxes on capital income. Furthermore, Pareto-efficient taxes on capital income tend to be higher for scale-dependent returns than for type-dependent returns. This underscores the need for additional empirical research to better understand the nature of return heterogeneity.

Future theoretical research may extend the current paper in a number of directions. First, it would be interesting to consider a setting in which there is no one-to-one mapping between labour income and capital income. This would imply that capital income is heterogeneous even among people with the same labour income. This complicates the search for Pareto-improving reforms of the tax mix. At the same time, it may strengthen the case for taxes on capital income, as they would yield redistributional gains beyond what can be achieved with a tax on labour income alone.

Second, it would be interesting to add idiosyncratic and systematic risk and portfolio choice to analyse optimal taxes on capital income with heterogeneous returns. If risk aversion is correlated with earnings ability, it might be optimal to distort risk-taking behaviour for income redistribution by differentiating taxes on safe and risky assets.

Third, this paper's two-period model structure might be extended to allow for a multiple-period life-cycle or overlapping-generations model to explore the quantitative robustness of our results in more realistic multiple-period models.

Finally, the model could be extended with entrepreneurial effort, tax avoidance in capital taxes and cross-border shopping. Such settings would allow for an expansion of the government instrument set to include wealth and consumption taxes.

Erasmus University Rotterdam, The Netherlands
Vrije Universiteit Amsterdam, The Netherlands
Erasmus University Rotterdam, The Netherlands
Organisation for Economic Co-operation & Development, France

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Agarwal, S., Driscoll, J.C., Gabaix, X. and Laibson, D. (2009). 'The age of reason: Financial decisions over the life cycle and implications for regulation', *Brookings Papers on Economic Activity*, vol. 2009(2), pp. 51–117.

Aiyagari, S.R. (1995). 'Optimal capital income taxation with incomplete markets, borrowing constraints and constant discounting', *Journal of Political Economy*, vol. 103(6), pp. 1158–75.

Atkinson, A.B., Piketty, T. and Saez, E. (2011). 'Top incomes in the long run of history', *Journal of Economic Literature*, vol. 49(1), pp. 3–71.

Atkinson, A.B. and Sandmo, A. (1980). 'Welfare implications of the taxation of savings', Economic Journal, vol. 90(359), pp. 529–49.

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

Attanasio, O.P. and Weber, G. (2010). 'Consumption and saving: Models of intertemporal allocation and their implications for public policy', *Journal of Economic Literature*, vol. 48(3), pp. 693–751.

Bach, L., Calvet, L.E. and Sodini, P. (2020). 'Rich pickings? Risk, return, and skill in household wealth', *American Economic Review*, vol. 110(9), pp. 2703–47.

Banks, J. and Diamond, P.A. (2010). 'The base for direct taxation', in (J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba, eds.), *Dimensions of Tax: The Mirrlees Review*, pp. 548–648, Oxford: Oxford University Press.

Barber, B.M., Odean, T. and Zheng, L. (2005). 'Out of sight, out of mind: The effects of expenses on mutual fund flows', *Journal of Business*, vol. 78(6), pp. 2095–120.

Bastani, S., Karlsson, K. and Waldenström, D. (2023). 'The different returns to cognitive ability in the labor and capital markets', Working Paper 2023:8, IFAU.

Bastani, S. and Waldenström, D. (2020). 'How should capital be taxed?', *Journal of Economic Surveys*, vol. 34(4), pp. 812–46.

Benartzi, S. and Thaler, R.H. (2001). 'Naive diversification strategies in defined contribution saving plans', *American Economic Review*, vol. 91(1), pp. 79–98.

Benhabib, J., Bisin, A. and Luo, M. (2019). 'Wealth distribution and social mobility in the US: A quantitative approach', *American Economic Review*, vol. 109(5), pp. 1623–47.

Best, M.C., Cloyne, J.S., Ilzetzki, E. and Kleven, H.J. (2020). 'Estimating the elasticity of intertemporal substitution using mortgage notches', *Review of Economic Studies*, vol. 87(2), pp. 656–90.

Black, S.E., Devereux, P.J., Landaud, F. and Salvanes, K.G. (2023). 'Where does wealth come from? Measuring lifetime resources in Norway', *Journal of Economic Perspectives*, vol. 37(4), pp. 115–36.

Blundell, R. and MaCurdy, T. (1999). 'Labor supply: A review of alternative approaches', in (O.C. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, pp. 1559–695, Amsterdam: Elsevier.

Boadway, R. and Spiritus, K. (2024). 'Optimal taxation of normal and excess returns to risky assets', *Scandinavian Journal of Economics*, published online ahead of print.

Calvet, L.E., Campbell, J.Y. and Sodini, P. (2007). 'Down or out: Assessing the welfare costs of household investment mistakes', *Journal of Political Economy*, vol. 115(5), pp. 707–47.

Campbell, J.Y. (2016). 'Restoring rational choice: The challenge of consumer financial regulation', *American Economic Review: Papers & Proceedings*, vol. 106(5), pp. 1–30.

Chamley, C. (1986). 'Optimal taxation of capital income in general equilibrium with infinite lives', *Econometrica*, vol. 54(3), pp. 607–22.

Choi, J.J., Laibson, D. and Madrian, B.C. (2005). 'Are empowerment and education enough? Underdiversification in 401 (k) plans', *Brookings Papers on Economic Activity*, vol. 2005(2), pp. 151–213.

Choi, J.J., Laibson, D. and Madrian, B.C. (2010). 'Why does the law of one price fail? An experiment on index mutual funds', *Review of Financial Studies*, vol. 23(4), pp. 1405–32.

Choi, J.J., Laibson, D. and Madrian, B.C. (2011). '$100 Bills on the sidewalk: Suboptimal investment in 401 (k) plans', *Review of Economics and Statistics*, vol. 93(3), pp. 748–63.

Christiansen, V. (1981). 'Evaluation of public projects under optimal taxation', *Review of Economic Studies*, vol. 48(3), pp. 447–57.

Christiansen, V. (1984). 'Which commodity taxes should supplement the income tax?', *Journal of Public Economics*, vol. 24(2), pp. 195–220.

Christiansen, V. (1993). 'A normative analysis of capital income taxes in the presence of aggregate risk', *Geneva Papers on Risk and Insurance Theory*, vol. 18(1), pp. 55–76.

Christiansen, V. and Tuomala, M. (2008). 'On taxing capital income with income shifting', *International Tax and Public Finance*, vol. 15(4), pp. 527–45.

Conesa, J.C., Kitao, S. and Krueger, D. (2009). 'Taxing capital? Not a bad idea after all!', *American Economic Review*, vol. 99(1), pp. 25–48.

Corlett, W.J. and Hague, D.C. (1953). 'Complementarity and the excess burden of taxation', *Review of Economic Studies*, vol. 21(1), pp. 21–30.

Cremer, H., Pestieau, P. and Rochet, J.C. (2001). 'Direct versus indirect taxation: The design of tax structure revisited', *International Economic Review*, vol. 42(3), pp. 781–99.

Diamond, P.A. (1998). 'Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates', *American Economic Review*, vol. 88(1), pp. 83–95.

Diamond, P.A. and Mirrlees, J.A. (1978). 'A model of social insurance with variable retirement', *Journal of Public Economics*, vol. 10(3), pp. 295–336.

Diamond, P.A. and Saez, E. (2011). 'The case for a progressive tax: From basic research to policy recommendations', *Journal of Economic Perspectives*, vol. 25(4), pp. 165–90.

Diamond, P.A. and Spinnewijn, J. (2011). 'Capital income taxes with heterogeneous discount rates', *American Economic Journal: Economic Policy*, vol. 3(4), pp. 52–76.

Erosa, A. and Gervais, M. (2002). 'Optimal taxation in life-cycle economies', *Journal of Economic Theory*, vol. 105(2), pp. 338–69.

Fagereng, A., Guiso, L., Malacrino, D. and Pistaferri, L. (2020). 'Heterogeneity and persistence in returns to wealth', *Econometrica*, vol. 88(1), pp. 115–70.

Farhi, E., Sleet, C., Werning, I. and Yeltekin, S. (2012). 'Non-linear capital taxation without commitment', *Review of Economic Studies*, vol. 79(4), pp. 1469–93.

Federal Reserve Bank of St. Louis. (2020). *30-Year Treasury Constant Maturity Rate, Percent, Annual, Not Seasonally Adjusted*, St. Louis, MO: Federal Reserve Bank of St. Louis.

Feenberg, D. and Coutts, E. (1993). 'An introduction to the TAXSIM model', *Journal of Policy Analysis and Management*, vol. 12(1), pp. 189–94.
Ferey, A., Lockwood, B. and Taubinsky, D. (2024).'Sufficient statistics for nonlinear tax systems with general across-income heterogeneity', *American Economic Review*, vol. 114(10), pp. 3206–49.
Gabaix, X., Lasry, J.M., Lions, P.L. and Moll, B. (2016). 'The dynamics of inequality', *Econometrica*, vol. 84(6), pp. 2071–111.
Gahvari, F. and Micheletto, L. (2016). 'Capital income taxation and the Atkinson–Stiglitz theorem', *Economics Letters*, vol. 147, pp. 86–9.
Gerritsen, A. (2016). 'Optimal taxation when people do not maximize well-being', *Journal of Public Economics*, vol. 144, pp. 122–39.
Gerritsen, A. (2024). 'Optimal nonlinear taxation: A simpler approach', *International Tax and Public Finance*, vol. 31(2), pp. 486–510.
Goetzmann, W.N. and Kumar, A. (2008). 'Equity portfolio diversification', *Review of Finance*, vol. 12(3), pp. 433–63.
Golosov, M., Kocherlakota, N. and Tsyvinski, A. (2003). 'Optimal indirect and capital taxation', *Review of Economic Studies*, vol. 70(3), pp. 569–87.
Golosov, M., Troshkin, M., Tsyvinski, A. and Weinzierl, M. (2013). 'Preference heterogeneity and optimal capital income taxation', *Journal of Public Economics*, vol. 97, pp. 160–75.
Golosov, M., Tsyvinski, A. and Werquin, N. (2014). 'A variational approach to the analysis of tax systems', Working Paper 20780, NBER.
Gordon, R.H. (1985). 'Taxation of corporate capital income: Tax revenues versus tax distortions', *Quarterly Journal of Economics*, vol. 100(1), pp. 1–27.
Gordon, R.H. and Kopczuk, W. (2014). 'The choice of the personal income tax base', *Journal of Public Economics*, vol. 118, pp. 97–110.
Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S. and Chen, D. (2023). 'Use it or lose it: Efficiency and redistributional effects of wealth taxation', *Quarterly Journal of Economics*, vol. 138(2), pp. 835–94.
Havránek, T. (2015). 'Measuring intertemporal substitution: The importance of method choices and selective reporting', *Journal of the European Economic Association*, vol. 13(6), pp. 1180–204.
Hellwig, C. and Werquin, N. (2024). 'Using consumption data to derive optimal income and capital tax rates', Ms., Toulouse School of Economics.
Hubbard, R.G. and Judd, K.L. (1986). 'Liquidity constraints, fiscal policy, and consumption', *Brookings Papers on Economic Activity*, vol. 1986(1), pp. 1–59.
Immervoll, H., Kleven, H.J., Kreiner, C.T. and Saez, E. (2007). 'Welfare reform in European countries: A microsimulation analysis', Economic Journal, vol. 117(516), pp. 1–44.
Jacobs, B. (2013). 'From optimal tax theory to applied tax policy', *FinanzArchiv*, vol. 69(3), pp. 338–89.
Jacobs, B. (2018). 'The marginal cost of public funds is one at the optimal tax system', *International Tax and Public Finance*, vol. 24(4), pp. 883–912.
Jacobs, B. and Boadway, R. (2014). 'Optimal linear commodity taxation under optimal non-linear income taxation', *Journal of Public Economics*, vol. 117(1), pp. 201–10.
Jacobs, B. and Bovenberg, L.A. (2010). 'Human capital and optimal positive taxation of capital income', *International Tax and Public Finance*, vol. 17(5), pp. 451–78.
Jacobs, B. and Rusu, A.V. (2018). 'Why is the long-run tax on capital income zero? Explaining the Chamley–Judd result', Ms., Erasmus University Rotterdam.
Jacobs, B. and Schindler, D. (2012). 'On the desirability of taxing capital income in optimal social insurance', *Journal of Public Economics*, vol. 96(9–10), pp. 853–68.
Jacquet, L. and Lehmann, E. (2021). 'Optimal income taxation with composition effects', *Journal of the European Economic Association*, vol. 19(2), pp. 1299–341.
Jacquet, L., Lehmann, E. and Van der Linden, B. (2013). 'Optimal redistributive taxation with both intensive and extensive responses', *Journal of Economic Theory*, vol. 148(5), pp. 1770–805.
Jakobsen, K., Jakobsen, K., Kleven, H. and Zucman, G. (2020). 'Wealth taxation and wealth accumulation: Theory and evidence from Denmark', *Quarterly Journal of Economics*, vol. 135(1), pp. 329–88.
Judd, K.L. (1985). 'Redistributive taxation in a simple perfect foresight model', *Journal of Public Economics*, vol. 28(1), pp. 59–83.
Kaymak, B., Leung, D. and Poschke, M. (2022). 'Accounting for wealth concentration in the United States', Working Paper 22-28, Federal Reserve Bank of Cleveland.
King, M.A. (1980). 'Savings and taxation', in (G.A. Hughes and G.M. Heal, eds.), *Public Policy and the Tax System*, pp. 1–35, London: Allen and Unwin.
Koehne, S. and Sachs, D. (2022). 'Pareto-improving reforms of tax deductions', *European Economic Review*, vol. 148, article ID 104214.
Kristjánsson, A.S. (2016). 'Optimal taxation with endogenous return to capital', Ms., Oslo University.
Lusardi, A., Michaud, P.C. and Mitchell, O.S. (2017). 'Optimal financial knowledge and wealth inequality', *Journal of Political Economy*, vol. 125(2), pp. 431–77.
Lusardi, A. and Mitchell, O.S. (2011). 'Financial literacy around the world: An overview', *Journal of Pension Economics and Finance*, vol. 10(4), pp. 497–508.

Meghir, C. and Phillips, D. (2010). 'Labour supply and taxes', in (J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba, eds.), *Dimensions of Tax Design: The Mirrlees Review*, pp. 202–74, Oxford: Oxford University Press.
Mill, J.S. (1848). *Principles of Political Economy*, London: John W. Parker.
Mirrlees, J.A. (1971). 'An exploration in the theory of optimum income taxation', *Review of Economic Studies*, vol. 38(2), pp. 175–208.
Mirrlees, J.A. (1976). 'Optimal tax theory: A synthesis', *Journal of Public Economics*, vol. 6(4), pp. 327–58.
Ordover, J.A. and Phelps, E.S. (1979). 'The concept of optimal taxation in the overlapping–generations model of capital and wealth', *Journal of Public Economics*, vol. 12(1), pp. 1–26.
Pigou, A.C. (1928). *A Study in Public Finance*, London: Macmillan.
Piketty, T. (2014). *Capital in the Twenty-First Century*, Cambridge, MA: Harvard University Press.
Reis, C. (2010). 'Entrepreneurial labor and capital taxation', *Macroeconomic Dynamics*, vol. 15(3), pp. 326–35.
Saez, E. (2001). 'Using elasticities to derive optimal income tax rates', *Review of Economic Studies*, vol. 68(1), pp. 205–29.
Saez, E. (2002). 'The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes', *Journal of Public Economics*, vol. 83(2), pp. 217–30.
Saez, E. and Stantcheva, S. (2018). 'A simpler theory of optimal capital taxation', *Journal of Public Economics*, vol. 162, pp. 120–42.
Saez, E. and Zucman, G. (2016). 'Wealth inequality in the United States since 1913: Evidence from capitalized income tax data', *Quarterly Journal of Economics*, vol. 131(2), pp. 519–78.
Saez, E. and Zucman, G. (2019). 'Progressive wealth taxation', *Brookings Papers on Economic Activity*, Fall, pp. 437–533.
Scheuer, F. and Slemrod, J. (2021). 'Taxing our wealth', *Journal of Economic Perspectives*, vol. 35(1), pp. 207–30.
Scheuer, F. and Werning, I. (2017). 'The taxation of superstars', *Quarterly Journal of Economics*, vol. 132(1), pp. 211–70.
Scheuer, F. and Wolitzky, A. (2016). 'Capital taxation under political constraints', *American Economic Review*, vol. 106(8), pp. 2304–28.
Schindler, D. (2008). 'Taxing risky capital income—a commodity taxation approach', *FinanzArchiv*, vol. 64(3), pp. 311–33.
Schulz, K. (2023). 'Redistribution of return inequality', Ms., University of Mannheim.
Seim, D. (2017). 'Behavioral responses to wealth taxes: Evidence from Sweden', *American Economic Journal: Economic Policy*, vol. 9(4), pp. 395–421.
Shourideh, A. (2014). 'Optimal taxation of wealthy individuals', Ms., Wharton School, University of Pennsylvania.
Spiritus, K., Lehmann, E., Renes, S. and Zoutman, F.T. (forthcoming). 'Optimal taxation with multiple incomes and types', *Theoretical Economics*.
Stiglitz, J.E. (1982). 'Self-selection and Pareto efficient taxation', *Journal of Public Economics*, vol. 17(2), pp. 213–40.
Stiglitz, J.E. (1985). 'Inequality and capital taxation', Technical report 457, Stanford University.
Stiglitz, J.E. (2000). *Economics of the Public Sector*, 3rd edn, New York: W.W. Norton.
Stiglitz, J.E. (2018). 'Pareto efficient taxation and expenditures: Pre- and re-distribution', *Journal of Public Economics*, vol. 162, pp. 101–19.
US Census Bureau. (2019). 'Current population census, annual social and economic supplement', https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-10.html (last accessed: 25 September 2024).
Van Rooij, M., Lusardi, A. and Alessie, R. (2011). 'Financial literacy and stock market participation', *Journal of Financial Economics*, vol. 101(2), pp. 449–72.
Varian, H.R. (1980). 'Redistributive taxation as social insurance', *Journal of Public Economics*, vol. 14(1), pp. 49–68.
Vermeulen, P. (2018). 'How fat is the top tail of the wealth distribution?', *Review of Income and Wealth*, vol. 64(2), pp. 357–87.
Von Gaudecker, H.M. (2015). 'How does household portfolio diversification vary with financial literacy and financial advice?', *Journal of Finance*, vol. 70(2), pp. 489–507.
Werning, I. (2007). 'Optimal fiscal policy with redistribution', *Quarterly Journal of Economics*, vol. 122(3), pp. 925–67.
Yitzhaki, S. (1987). 'The relation between return and income', *Quarterly Journal of Economics*, vol. 102(1), pp. 77–95.
Zoutman, F.T. (2018). 'The effect of capital income taxation on household savings', Ms., Norwegian School of Economics.

The Economic Journal, **135** (*January*), 180–211 https://doi.org/10.1093/ej/ueae083 © The Author(s) 2024. Published by Oxford University Press on behalf of Royal Economic Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.
Advance Access Publication Date: 16 September 2024