July 10th 2024

ISSN 1011-8888

Institute of economic studies

WORKING PAPERS

WP24:03

July 2024

Optimal Taxation with Endogenous Return on Capital

Arnaldur Sölvi Kristjánsson

Mail address:

Faculty of Economics

University of Iceland

Oddi at Sturlugata,

102 Reykjavik,

Email: askristjansson@hi.is

Optimal Taxation with Endogenous Return on Capital*

Arnaldur Sölvi Kristjánsson†

July 2024

Abstract

This paper characterises optimal taxation when rates of return are affected by effort, ability, and financial advice. When the government observes wealth and capital income, the optimal marginal tax rate on capital income is positive, whereas the rate on wealth is negative in the baseline model. When wealth is not observed, the optimal marginal tax rate on capital income remains positive. If inequality in labour market productivity is sufficiently large compared to investment ability, the marginal tax rate on labour income exceeds the rate on capital income, and vice versa.

Keywords: Optimal taxation, capital taxation, endogenous return on capital

JEL Codes: G11, H21, H24

*I would like to thank Vidar Christiansen, Ray Rees, Agnar Sandmo, Sören Blomquist, Etienne Lehmann, Jacopo Bizzotto and Karl O. Moene for useful suggestions and comments.

†Department of Economics, University of Iceland. E-mail: askristjansson@hi.is

1 Introduction

The taxation of capital versus labour income has been the subject of academic debate for many years. Increased income and wealth concentration has attracted attention among academics regarding which tax instruments should be used. Debates focus on both the relative importance of labour and capital taxation, and how capital should be taxed, especially the use of capital income taxes versus wealth taxes.

In this paper, I study optimal taxation of labour income, wealth and capital income in a model where return on savings depends on effort, ability, and financial advice. The government has a utilitarian welfare function, and therefore redistribution from less-skilled to high-skilled individuals is desirable. Optimal taxation involves both the taxation of labour income and capital income. Because more skilled individuals are, conditional on income, more willing to save, exert more investment effort and spend more money on financial advice, capital income provides information on the underlying skill level and should therefore be taxed.

Comparing a capital income tax with a wealth tax that generates equivalent revenues, the former levies a heavier tax on excess returns (see Bastani & Waldenström, 2023). When returns depend on investment sophistication, which depends on the utilisation of time and financial resources, the tax system should emphasise taxing capital income over wealth, as capital income provides more information on the underlying skill level. Since the government aims to redistribute from the skilled to the low-skilled, optimal tax policy emphasises taxing capital income.

Marginal tax rates on labour income should exceed those on capital income when differences among individuals are predominantly due to labour market productivity. The reverse is true when differences primarily arise from investment ability. Thus, the government should impose larger marginal tax rates on the income type that is the predominant source of inequality.

The assumption that everyone faces a homogeneous interest rate is unrealistic. Broadly speaking, there are two main explanations for why rates of return are heterogeneous. The first explanation is that information on which investment options are likely to be good is asymmetric. This asymmetry exists because acquiring information on which investment options are good involves costs, such as the time needed for market research and the resources required to collect information. Just as individuals differ in their labour market productivity, they may also differ in their ability to increase investment sophistication. Secondly, differential rates of return arise due to uncertainty, an inherent feature of financial markets. This paper deals only with the former effect.

A growing literature, which is reviewed in the next section, provides an overview of the extent to which return heterogeneity can be explained by persistent differences in returns, where returns increase with ability (so-called *type-dependency*) and to what extent they increase with wealth, which may reflect fixed costs in wealth management (so-called *scale-dependency*). In general, risk does not fully explain return heterogeneity. Both type- and scale-dependency contribute to return heterogeneity, though there is no consensus on the relative importance of these factors (see Bach et al., 2020; Fagereng et al., 2020).

Numerous studies have analysed optimal taxation of capital with return heterogeneity, with both type- and scale-dependency. The studies most closely related to ours are Gahvari & Micheletto (2016) and Gerritsen et al. (2020), while other studies also address the subject (Gaillard & Wangner, 2021; Guvenen et al., 2023; Schulz, 2021; Zanoutene, 2023). See further discussion in Section 2.2. Although mechanisms differ between studies, the general conclusion is that return heterogeneity provides a rationale for higher taxes on capital. Except for Guvenen et al. (2023) and Zanoutene (2023), the studies primarily consider one type of capital taxation, mostly capital income taxation.

My main contribution is to consider a model with type-dependent returns, analysing different tax instruments and various informational assumptions for the government, in a model that allows for a comparison of marginal tax rates between the tax instruments.

I set up a two-period model of savings where individuals exert investment effort, which increases the rate of return, and they work in the labour market. Individuals differ in investment ability and labour market productivity. The government aims to redistribute resources from the skilled to the less skilled. I analyse whether the existence of different investment abilities and the possibility of exerting investment effort provide a rationale for taxing capital. Preferences are separable between consumption and leisure, satisfying the Atkinson & Stiglitz (1976) theorem, which states that the optimal capital income tax is zero under this condition. I show that a zero tax on capital is not optimal when one introduces the possibility of exerting investment effort, and individuals differ either in labour market productivity or in investment abilities.

There are two informational assumptions. In the first scenario, the government observes savings and capital income. In this scenario, the government aims to distort the investment decision while keeping the savings decision undistorted. This can be achieved through a positive marginal tax on capital income and a negative marginal tax on wealth. The choice of investment effort and labour supply depends on skill since, conditional on income, more skilled individuals enjoy more leisure. More skilled individuals are therefore more willing to increase their investment effort. Therefore, it is efficient to condition taxation on capital income to redistribute resources. That is, capital income provides the government with information on an individual's underlying skill level and should be used for taxation. However, savings do not depend on skill levels and thus do not help distinguish individuals with different skill levels.

In the second scenario, savings are not observed by tax authorities, whereas capital income is observed. In this scenario, savings are distorted by the capital income tax. The rationale for taxing capital income remains similar to that in the first scenario. Conditional on income, more skilled individuals have a higher rate of return and a greater marginal propensity to save. Thus, capital income depends on an individual's underlying skill level and should therefore be taxed. Since savings are not observed, the tax system cannot distort investment effort while keeping savings undistorted.

In an extension where individuals can spend money on financial advice and the government observes savings and capital income, I show that the optimal tax system ensures the intertemporal allocation is distorted downwards. Regarding the marginal tax rates on wealth and capital income, the emphasis is on the capital income tax, though the signs are generally ambiguous.

The remainder of the paper is organised as follows. In the next section, I review relevant empirical literature on return heterogeneity and related theoretical studies on optimal capital taxation. Section 3 introduces the model and presents the first-best allocation. In Section 4, I present optimality conditions when the government observes savings. The following section presents optimality conditions when savings are not observed. Section 6 presents optimality conditions when individuals have access to financial advice. Finally, Section 7 concludes.

2 Related Literature

2.1 Empirical Literature

The fact that rates of return are heterogeneous is indisputable. However, why this is the case and to what degree it is related to uncertainty and risk remains less clear. Several related literatures analyse differences in rates of return, providing some insights into the influences of risk and other factors.

First, the most direct evidence of heterogeneous returns comes from papers that directly analyse return heterogeneity. In an early contribution, Yitzhaki (1987) finds that returns increase with income, most likely due to differences in risk aversion. Piketty (2014) finds that returns on the endowments of US universities systematically increase with the size of the endowment. Saez & Zucman (2016) show that the same pattern emerges for the universe of U.S. foundations.

Two recent papers using Scandinavian administrative panel data find considerable return heterogeneity. Fagereng et al. (2020) argue that persistency in return heterogeneity is partly driven by risk taking and asset scale, but also by other factors such as financial sophistication, financial information, and entrepreneurial talent.[1] Bach et al. (2020) analyse the relationship between returns and wealth using Swedish panel data. They decompose expected returns into a type and scale effect, using twin pair fixed effects and wealth percentiles as explanatory variables, the results indicate that both type and scale dependencies contribute to return heterogeneity.

Second, a literature on drivers of wealth inequality has established that labour income disparities cannot account for the wealth concentration observed in the data, particularly at the very top (see e.g., De Nardi & Fella, 2017). Gabaix et al. (2016), Benhabib et al. (2011), and Xavier (2021) show that return heterogeneity can generate the wealth inequality observed in the data. Gabaix et al. (2016) suggest that both type- and scale-dependency in returns for wealth inequality and its increase over time.

Third, a literature on household finance reveals the importance of financial literacy on financial outcomes. This literature has established a positive association between financial literacy and significant outcomes such as planning for retirement, participating in the stock market, diversifying investments, and avoiding high-cost borrowing.[2] Interpreting this association as causal is challenging; however, both instrumental variables and experimental approaches suggest that literacy causally influences financial outcomes—and not vice versa (see e.g., Lusardi & Mitchell, 2014). Lusardi et al. (2017) find that 30-40% of retirement inequality can be attributed to differences in financial knowledge.

Some level of financial ignorance and less sophistication in investment strategies may be optimal, as there are costs associated with increasing financial literacy (Campbell, 2006; Lusardi & Mitchell, 2014). Due to transaction and search costs, individuals may rationally choose to under-diversify, as they weigh the costs against the benefits (see review in Guiso & Jappelli, 2008).

Analysing returns on savings accounts, virtually riskless assets, Deuflhard et al. (2019) find considerable heterogeneity in the Netherlands. The results suggest that a lack of information explains why households do not choose options with the highest returns, implying that financial literacy significantly affects the ability to identify high-return accounts.[3]

[1]The authors, for example, find that return heterogeneity in risk-free deposit accounts is partly explained by individual attributes and fixed effects, suggesting that some individuals are more adept at spotting high-return banks.

[2]For literature reviews, see Campbell (2006), Hastings et al. (2013), Lusardi & Mitchell (2014), and Stolper & Walter (2017).

[3]Bastani et al. (2023) find that cognitive ability is positively associated with returns, including for bank deposits.

2.2 Theoretical Literature

Well known results that capital income should be untaxed are from Atkinson & Stiglitz (1976), Chamley (1986) and Judd (1985). A result from a two-period extension of the Atkinson & Stiglitz (1976) model is that intertemporal allocations should not be distorted, i.e. capital income should not be taxed.[4] Chamley (1986) and Judd (1985) show that taxes on capital income are zero in the long run in an infinitely-lived representative-agent Ramsey model.[5]

The literature has identified numerous conditions under which the Atkinson-Stiglitz theorem does not hold. These include heterogeneous preferences, different initial wealth, the presence of income shifting, reducing distortions from labour income tax, uncertain future wages and borrowing constraints, taxing economic rents, and heterogeneous returns (see further below).[6]

In the model by Christiansen & Tuomala (2008) individuals can, at a cost, shift income from the labour income to the capital income tax base. In a number of papers which deal with human capital investment and optimal taxation, individuals spend time and/or resources on education to increase their future productivity (see e.g., Bovenberg & Jacobs, 2005; Jacobs & Bovenberg, 2010). In my model, individuals spend time on increasing future capital income.

My paper closely relates to the literature on optimal taxation with varying rates of returns. The general results are that return heterogeneity calls for a higher taxation on capital income. While mechanisms and appropriate tax tools differ between the models.

An early approach is the model by Stiglitz (1985), where individuals differ in the rate of return they obtain. Gahvari & Micheletto (2016) set up a closely related model to mine, where the return is an increasing function of labour market productivity and labour income.

Gerritsen et al. (2020) set a model featuring type- and scale-dependency, They find that capital income taxes are positive under both heterogeneities—averaging around 10% for type-dependent and 25% for scale-dependent returns. For type-dependent returns, taxing capital income targets previously earned income and ability rents. In the case of scale-dependent returns, taxing capital income later in life is efficient, reflecting the government's practice of taxing income upon earning.

Schulz (2021) presents a model with type- and scale-dependency in returns. Optimal capital income taxes are expressed in terms of sufficient statistics, and yield a standard inverse-elasticity rule condition. Scale dependency leads to a lower or unchanged optimal

[4]See Stiglitz (1985, 1987) for the extension of the Atkinson-Stiglitz model.

[5]However, Straub & Werning (2020) show that when the intertemporal elasticity of substitution is less than one, optimal capital income tax rates can be positive.

[6]For a thorough literature review and further references on how and why capital should be taxed, see Bastani & Waldenström (2020), Banks & Diamond (2010), and Jacobs (2013).

capital income tax rate, though scale decency increases inequality, calling for a higher tax, it increases the savings elasticity, calling for a lower tax rate.

Zanoutene (2023) presents a two period model with scale dependency and risk, where higher savings and luck can increase returns. Optimal capital taxation serves as an insurance device against risky returns. He argues that scale dependency alone, without risk, is not sufficient to justify capital income taxation.

Gaillard & Wangner (2021) analyse a wealth tax with heterogeneous returns. When heterogeneity is due to scale-dependency, rent extraction motives are absent and the optimal wealth tax rate is negative. When heterogeneity is due to type-dependency, the optimal wealth tax rate is positive.

Finally, Guvenen et al. (2023) set up a model quantitative overlapping-generation model with heterogeneous returns due to heterogeneous entrepreneurial productivity and incomplete financial markets. They present optimal linear tax rates on capital income and wealth. They show that a shift from capital income to wealth taxation to be welfare improving, as this would shift the tax burden from the more productive to less productive entrepreneurs, ensuring a more efficient allocation of capital.

Boadway & Spiritus (2021) considers a model with risky returns. Optimal taxation on excess returns is positive and only serves an insurance role, while the optimal rate on risk-free returns serves a redistributional role if returns reveal information about investor types. Finally, Saez & Stantcheva (2018) present optimal tax formulas for capital income that allow for return heterogeneity.

3 The Model

Following Mirrlees (1971), optimal income tax models treat different observed incomes as outcomes of exogenously given abilities and endogenously determined labour supply. In my model, individuals differ not only in their productivity for paid work, denoted by w, but also in investment abilities, denoted by a. Individuals have two ways to increase their resources. First, through standard labour supply, where individuals work in the labour market and earn wages. Second, individuals can spend time managing their portfolio, which raises their return on savings.

Labour market productivity and investment ability, both exogenously given, are innate traits that individuals cannot alter. In all other respects, individuals are identical.

Individuals live for two periods. They work in the first period and consume in both periods. Period one can be thought of as working years and period two as retirement years. In the first period, they have to decide how much time to spend on the labour market and how much time to devote to investments. Time spent working in the labour market is denoted by L and time spent investing is denoted by E. The second decision individuals have to make is how much to save in the first period. Thereby, individuals decide how to split consumption between the two periods.

Agents supply L units of labour in the first period. The labour market is perfectly competitive, and individuals with different productivity levels are perfect substitutes, so workers receive a wage of w based on their exogenously given abilities. Labour income is denoted as $Y = wL$.

Following the literature on financial literacy, time spent investing increases the return to savings.[7] This can be considered as investment in financial literacy, increasing financial sophistication. This helps individuals to invest in higher-return assets and reduce investment expenses.

The economy is small and open, with individuals engaging in investments within an international investment market. As the economy is small and open, the behaviour of individuals have no general equilibrium effects.

Capital income is denoted by $k(E, s, a)$ which increases with E, s and a, where s denotes savings. Furthermore, it possesses the following properties:

$$k_{sa}, k_{sE}, k_{Ea} > 0, k_{ss}, k_{EE} \leq 0.$$

The cross-derivative $k_{sa} > 0$ indicates that more able investors get a higher rate of return and $k_{Ea} > 0$ implies that more able investors are more efficient, aligning with my definition of a more able investor. The positive cross-derivative k_{sE} indicates that the rate of return increases with investment effort.

There is a given amount of heterogeneous investment options. If each investment option is of finite size, then for a given E and a an increase in s should either decrease or keep k_s unchanged. In reality, interest rates sometimes increase with savings. For example, some investment options require a minimum investment amount. In the model, decreasing returns to scale in savings is a sufficient condition.

The budget constraint of individuals in periods 1 and 2 are

$$\begin{aligned} c_1 =& wL - s - t, \\ c_2 =& s + k(E, a, s) - T, \end{aligned}$$

[7] Delavande et al. (2008) presents a two-period model with portfolio allocation across safe and risky assets. In addition, individuals invest in financial sophistication using time and monetary resources. Increased financial sophistication raises the risk-adjusted rate of return. Similar approaches can be found in Arrow (1987), Jappelli & Padula (2017), and Lusardi et al. (2017). In Section 6, individuals can also spend money on financial advice to gain information that will increase the rate of return.

where c_1 and c_2 are consumption in period 1 and 2, respectively, and t and T are tax payments in period 1 and 2, respectively.

Individuals have identical, separable, and additive utility functions, which are increasing and concave in c_1, c_2, and leisure. The separability between leisure and consumption aims to avoid the effects of complementarity in optimal taxation, an issue that has received significant attention in the literature (see Christiansen, 1984). The total time available is normalized to 1. Individuals face a time constraint such that leisure equals $1 - L - E$. The utility function is denoted by

$$U = u(c_1) + \psi(c_2) + v(1 - L - E), \tag{1}$$

where $u', \psi', v' > 0$ and $u'', \psi'', v'' < 0$.

The government has a utilitarian objective function, maximizing the sum of utilities. Given that individuals' utility functions are concave, the government has a redistributive motive. It is assumed that the government does not know individual skill level (w and a) or individual labour supply and investment effort (L and E). However, the government can observe both labour income and capital income at the individual level and knows the distribution of w and a along with individual preferences. I will analyse both the case when the government also observes s and when they don't. The former constitutes the case when the government observes capital income as well as wealth. The latter case is where the government only observes capital income and not wealth.

Unobservability of s is based on the notion that governments can conveniently observe the income stream from capital but not the stock of capital itself. Capital income, usually coming in the form of transactions (e.g., dividends and interest), is therefore comparatively simpler for tax authorities to observe. On the other hand, estimating the market value of assets is a much more daunting exercise.[8] I believe that both informational assumptions represent two extreme versions of reality, and therefore I analyse both cases.

The problem is solved using the direct approach, where the government assigns quantities (also called bundles) of pre- and post-tax incomes to every type in both periods. Then, the government solves the problem subject to the incentive constraints to prevent a certain type from choosing a bundle intended for another type (i.e., mimicking another type).

I will consider a discrete-type version of the Mirrlees model in the spirit of Stern (1982) and Stiglitz (1982). There are two dimensions, resulting in a total of four types of individuals. Labour market productivity can be either high or low, denoted by w^h and w^l, respectively,

[8] Slemrod & Gillitzer (2014) argue that basing tax liability on market transactions has several advantages. They say, "taxing capital gains on a realization basis rather than the theoretically preferable accrual basis takes advantage of the measurement advantage of market transactions. In contrast, estate and wealth taxation cannot, in general, take advantage of market transaction to reliably value wealth." (2014: 103).

with $w^h > w^l$. Similarly, investment ability can be either high or low, denoted by a^h and a^l, respectively, with $a^h > a^l$.[9]

3.1 First Best

In the first best, the government knows the investment ability and labour market productivity of all individuals, and is therefore not concerned with an incentive constraint. The government's problem is to maximise the sum of utilities

$$\max_{\left\{Y^i, B_1^i, s^i, E^i, B_2^i\right\}} \sum_{i}^{I=4} n^i \left[u(B_1^i - s^i) + \psi(B_2^i + s^i) + v\left(1 - Y^i/w^i - E^i\right)\right], \tag{2}$$

where n^i denotes the number of individuals of type i, $B_1^i = Y^i - t^i$ and $B_2^i = k(E^i, s^i, a^i) - T^i$ denote post-tax income in the first and second periods, respectively. In (2), the variables c_1, c_2 and L have been substituted for Y and s, a procedure that will be followed from now on.

The government has a certain revenue requirement in periods 1 and 2, denoted by g_1 and g_2, respectively. This can be interpreted as the required revenue for essential public goods. For simplicity, the government is restricted from borrowing or saving across these periods. However, this restriction does not impact the main findings.[10]

The budgetary and resource constraints faced by the government in periods 1 and 2 are, respectively,

$$\sum_{i=1}^{I=4} n^i (Y^i - B_1^i) \geq g_1, \quad \sum_{i=1}^{I=4} n^i (k(E^i, a^i, s^i) - B_2^i) \geq g_2. \tag{3}$$

The necessary conditions are

$$\frac{v_i'}{u_i'} \frac{1}{w^i} = MRS_Y^i = \left.\frac{dc_1^i}{dY^i}\right|_{\overline{U}} = 1, \tag{4a}$$

$$\frac{v_i'}{\psi_i'} \frac{1}{k_E^i} = MRS_K^i = \left.\frac{dc_2^i}{dK^i}\right|_{\overline{U}} = 1, \tag{4b}$$

[9] In the optimal tax literature, it is a standard assumption that individuals differ only in terms of labour market productivity. However, many models incorporate multiple dimensions of heterogeneity. See, for example Cremer et al. (2004) and Jacquet & Lehmann (2023).

[10] The inability to save or borrow, however, implies that the timing of taxation matters. In other words, Ricardian equivalence does not apply. If the government were permitted to borrow or save, on the other hand, Ricardian equivalence would be upheld. Including government borrowing/saving wouldn't alter the main results, but it would nonetheless influence the optimal intertemporal distribution. The impact of government borrowing primarily hinges on the interest rate the government is subjected to. Should the government face a high interest rate, it would undertake investments by levying substantial taxes in the first period and minimal taxes in the second period. Conversely, if faced with a low interest rate, the government would opt to borrow, allowing individuals to invest the borrowed funds.

$$\frac{u_i'}{\psi_i'} = MRS_c^i = -\frac{dc_2^i}{dc_1^i}\bigg|_{\overline{U}} = \frac{\lambda_1}{\lambda_2} = 1 + k_s^i. \tag{4c}$$

where MRS_Y^i denotes the marginal rate of substitution between labour income and present consumption, MRS_K^i denotes the marginal rate of substitution between capital income and future consumption, and MRS_c^i is the intertemporal marginal rate of substitution. The first two conditions show that the intratemporal marginal rates of substitution should equate to the marginal rates of transformation, which are 1. Condition (4c) shows that the intertemporal marginal rate of substitution should be equal to $1 + k_s^i$.

4 Government observes Wealth

The government observes individual capital income as well as individual savings but lacks information about individual abilities. The capital income assigned by the government is denoted by K, while $k(E, s, a)$ is capital income received by an individual. These two must be equal, i.e., $K = k(E, s, a)$.

To avoid difficulties with multidimensional screening, this section presents a two-type model with a perfect correlation between a and w, where $a^2 > a^1$ and $w^2 > w^1$ (a four-type model is examined in Section 5.3). The government's optimization problem is outlined in (2) with $I = 2$, subject to the revenue constraints (3) and the incentive constraint,

$$U^2 \geq \hat{U}^{21}, \tag{5}$$

where $\hat{U}^{21} = u(B_1^1 - s^1) + \psi(B_2^1 + s^1) + v(1 - Y^1/w^2 - \hat{E}^{21})$ is the utility of a type 2 individual choosing the bundle intended for a type 1 individual. Here, $\hat{E}^{21}$ is the investment effort made by a type 2 individual mimicking a type 1 individual, i.e. $K^1 = k(\hat{E}^{21}, s^1, a^2)$. This condition ensures that a type 2 individual does not choose the bundle designed for a type 1 individual.

Mimickers have more leisure time than the type they mimic, because they supply less labour due to their higher w, and they exert less investment effort due to their higher a, i.e., $L^1 + E^1 > \hat{L}^{21} + \hat{E}^{21}$.

The optimal allocation is characterised by:[11]

$$MRS_K^1 = 1 - \frac{\gamma\psi_1'}{n^1\lambda_2}\left[MRS_K^1 - M\hat{R}S_K^{21}\right] < 1, \tag{6a}$$

$$MRS_c^1 = 1 + k_s^1 = \frac{\lambda_1}{\lambda_2}, \tag{6b}$$

[11] In Appendix A, the Lagrangian is presented, the necessary conditions are derived and manipulated.

$$MRS_Y^1 = 1 - \frac{\gamma u_1'}{n^1 \lambda_1} \left[MRS_Y^1 - M\hat{R}S_Y^{21} \right] < 1, \tag{6c}$$

where γ is the Lagrange multiplier associated with the incentive constraint, λ_1 and λ_2 denote the Lagrange multipliers associated with the budget constraints for periods 1 and 2, respectively. $M\hat{R}S_K^{21} = \frac{\hat{v}_{21}'}{\psi_1'} \frac{1}{\hat{k}_E^{21}}$ and $M\hat{R}S_Y^{21} = \frac{\hat{v}_{21}'}{u_1'} \frac{1}{w^2}$ are the marginal rates of substitution for capital income and labour income, respectively, for mimickers (type 2 choosing the bundle intended for type 1).

At the optimum, type 2 individuals are undistorted. This follows the standard 'no distortion at the top' result as in the Mirrlees model, and will hold in all applications considered. Therefore, specific results are not shown.

Conditions (6a) and (6c) indicate that capital income and labour income should be distorted downwards, implying that investment effort and labour supply are distorted downwards. According to (6b), the intertemporal allocation, however, should remain undistorted, ensuring that the rate of return (k_s^i) will be constant across agents. Nonlinear taxation together with decreasing returns to scale ensure that rates of returns are equalised across individuals, ensuring efficient allocation of savings.

Type 2 individuals are, conditional on income, more inclined to work more for two reasons. First, they have more leisure, as previously discussed. Second, their higher wage rate, owing to greater productivity, results in a lower marginal rate of substitution (MRS_Y). Consequently, labour income provides the government with information on individual productivity. To redistribute from type 2 to type 1 individuals, the government should base taxes on labour income, thereby making mimicking less appealing and relaxing the binding incentive constraint.

The same applies to capital income. Type 2 individuals are, conditional on income, more willing to exert investment effort for two reasons: First, their leisure time is greater, as mentioned earlier. Second, their higher productivity as investors leads to a lower marginal rate of substitution (MRS_K). Therefore, capital income also informs the government about productivity levels. To redistribute from type 2 to type 1 individuals, capital income should be used for taxation purposes.

To demonstrate how conditions (6a)–(6c) can be implemented by a tax system, consider the individual's budget constraint with tax functions for both periods,[12]

$$\begin{aligned} c_1 &= Y - s - t(Y), \\ c_2 &= s + k(E, s, a) - T(s + k(E, s, a), k(E, s, a)). \end{aligned} \tag{7}$$

[12]Note that these tax functions are more restrictive than the allocations in (6).

Given conditions (6a) and (6c) alongside the individual's necessary conditions,[13] it follows that

$$t' > 0, \qquad T_K > 0, \qquad \text{and} \qquad T_W = -T_K \frac{k_s}{1+k_s} < 0. \tag{8}$$

A positive marginal tax on capital income (T_K) will distort investment effort and the intertemporal allocation. To keep the intertemporal allocation undistorted, wealth needs to be subsidised at the margin, meaning that the marginal subsidy of wealth (T_W) ensures the intertemporal allocation remains undistorted.

This is contrary to the results obtained in Guvenen et al. (2023), where the optimal capital income tax is negative, and the wealth tax positive. Such a system taxes the normal rate of return while favouring higher returns. In their model, a wealth tax serves to reallocate capital from individuals with low rates of returns to those with high rates of returns. In my model, higher returns reflect the underlying skill distribution and should be taxed rather than subsidised at the margin. Nonlinear taxes are designed to equalise returns, which is possible due to the decreasing returns to scale in savings. Distortions serve to tax those with high skills more heavily.

The above results are similar to those found in proposition 2 of Gerritsen et al. (2020), where capital income contains ability rents, when returns are type-dependent. Then a positive tax on capital income is optimal to ensure that these rents are taxed, which a tax on labour income does not tax. They also analyse optimal taxation in the presence of scale-dependency, then a capital income tax is due to alleviate the market failure by reallocating savings from those with low rates of returns to those with high. Taxing those with high returns later in life, as a tax on capital income does, allows them to better exploit their scale effects, since returns are increasing in savings. My model only includes type-dependency.

The magnitude of the marginal tax rates depends on the underlying inequalities, i.e. $w^2 - w^1$ and $a^2 - a^1$. If individuals differ only in terms of labour market productivity, then the marginal tax rate on labour income will exceed that on capital income, i.e. $t' > T_K$. Capital income and labour income provides the government with information on productivity because, conditional on income, more productive individuals are at the margin more willing to exert investment effort, and supply labour due to having more leisure. Therefore, the optimal marginal tax rates on labour income and capital income are positive. A further argument for distorting labour income is that more productive individuals are more willing to supply more labour due to their higher hourly wage rate. Therefore, labour income should be more heavily distorted.

The opposite is true when people only differ in terms of investment ability. Then, as

[13] The necessary conditions are: $MRS_K = 1 - T_K$, $MRS_Y = 1 - t'$, and $MRS_c = 1 + k_s(1 - T_K) - T_s$.

above, conditional on income, more able investors are at the margin more willing to exert investment effort and supply labour due to having more leisure. In addition, as they possess greater investment ability, the return to investment effort is higher, making them more inclined to exert investment effort.

In the optimal tax system, marginal tax rates on labour income should exceed those on capital income when differences among individuals are largely due to labour market productivity. The reverse is true when differences primarily arise from investment ability. Thus, the government should impose larger marginal tax rates on the income type that is the predominant source of inequality.

The results from this section are summarized in proposition 1.

Proposition 1 *In a tax system where the government observes savings, capital income, and labour income, optimal taxation involves a positive marginal tax rate on capital income and labour income, but a negative marginal tax rate on wealth. This will distort investment effort while keeping the intertemporal allocation undistorted. The marginal tax rate on labour income will exceed (fall short of) that on capital income if disparities among individuals are due solely to differences in labour market productivity (investment ability).*

5 Government does not observe Wealth

The government observes individual labour income and capital income but not savings, and knows the distribution of w and a as well as preferences. Before presenting the optimality conditions, comparative statics are presented.

5.1 Comparative Statics

The government offers bundles characterised by pre- and post-tax income in both periods, i.e., bundles in terms of (Y, B_1, K, B_2). Individuals select from these bundles offered by the government.

Upon selecting a bundle, individuals have no degree of freedom in terms of $L = Y/w$. For the bundle in the second period, individuals have one degree of freedom. They choose both E and s, but are constrained by the fact that capital income needs to equate to the level set by the government, or

$$k(E, a, s) = K, \tag{9}$$

where K is the quantity chosen by the government and $k(E, a, s)$ is capital income that individuals receive. Individuals choose E freely and let s adjust according to the constraint. This implicitly defines $s(E, a, K)$.

Now, an individual who chooses the bundle (Y, B_1, K, B_2) faces the problem

$$\max_{\{E\}} U = u(B_1 - s(E, a, K)) + \psi(B_2 + s(E, a, K)) + v(1 - Y/w - E).$$

This problem applies to all individuals, whether they are mimickers or not. The necessary condition is

$$U_E = -u'(B_1 - s(E, a, K))s_E + \psi'(B_2 + s(E, a, K))s_E - v'(1 - Y/w - E) = 0, \tag{10}$$

where $s_E = -k_E/k_s < 0$. Condition (10) indicates that an increase in E leads to an increase in first-period consumption (since savings are reduced, due to (9)), a decrease in second-period consumption (also due to (9)), and a decrease in leisure.

From the necessary conditions in (10), it follows that the optimal choice of E is a function of all the exogenous variables, i.e. $E^* = E(Y, B_1, K, B_2, w, a)$. To analyse the behaviour of mimickers, condition (10) is implicitly differentiated with respect to w and a,

$$\frac{dE}{dw} = \frac{-U_{Ew}}{U_{EE}} = \frac{v''}{-(u' - \psi')s_{EE} + (\psi'' + u'')s_E^2 + v''}\frac{Y}{w^2} > 0, \tag{11}$$

$$\frac{dE}{da} = \frac{-U_{Ea}}{U_{EE}} = \frac{(u' - \psi')s_{Ea} - (\psi'' + u'')s_E s_a}{-(u' - \psi')s_{EE} + (\psi'' + u'')s_E^2 + v''} < 0, \tag{12}$$

where $s_a = -k_a/k_s < 0$ and $s_{EE} = (k_E k_{sE} - k_{EE}k_s)/k_s^2 > 0$, hence $U_{EE} < 0$, and $s_{Ea} = 0$, see Appendix B. In (11) and (12), Y, B_1, K, B_2 and either a or w are held constant. Therefore, these derivatives indicate the behaviour of the mimicker. The first ratio on the RHS in (11) is less than one and since $\frac{dL}{dw} = -\frac{Y}{w^2}$, it follows that $\frac{dE}{dw} < -\frac{dL}{dw}$. This means that a mimicker with a higher w has more leisure than the less able worker. It follows from (12) that mimickers who have higher a has more leisure than the type they are mimicking.

To determine whether a mimicker saves more or less than the type being mimicked, $s(E^*, a, K)$ is implicitly differentiated,

$$\frac{ds}{dw} = s_E\frac{dE}{dw} < 0, \tag{13}$$

$$\frac{ds}{da} = s_E\frac{dE}{da} + s_a = \frac{-U_{Ea}s_E + s_a U_{EE}}{U_{EE}} = \frac{-(u' - \psi')s_{EE}s_a + v''}{U_{EE}} < 0. \tag{14}$$

When labour market productivity increases, individuals can achieve a given level of Y with a lower L. Consequently, the total utility of leisure increases. Given that the utility function is concave, individuals adjust by increasing E and decreasing s. As a result, a mimicker saves less and exerts more investment effort compared to the type being mimicked, i.e. $\hat{s}^{ji} < s^i$

and $\hat{E}^{ji} > E^i$, where $\hat{s}^{ji}$ and $\hat{E}^{ji}$ denote the savings and investment effort, respectively, of a type j mimicking a type i individual. With higher labour market productivity, a mimicker will mechanically supply less labour, i.e. $\hat{L}^{ji} < L^i$. Overall, a mimicker enjoys more leisure.

As investment ability increases, an individual requires less effort to invest and save to receive a given K. Therefore, it becomes beneficial for them to reduce both investment effort and savings. This leads high-ability mimickers to save less and exert lower investment effort compared to the types they mimic, i.e., $\hat{s}^{ji} < s^i$ and $\hat{E}^{ji} < E^i$.

The indirect utility function is denoted by $V(Y, B_1, K, B_2)$. The derivatives of the indirect utility function are derived from the envelope theorem,

$$\frac{\partial V}{\partial Y} = -\frac{v'}{w}, \qquad \frac{\partial V}{\partial B_1} = u', \qquad \frac{\partial V}{\partial K} = -\frac{u'}{k_s} + \frac{\psi'}{k_s} = -\frac{v'}{k_E}, \qquad \frac{\partial V}{\partial B_2} = \psi',$$

where $s_K = 1/k_s$ is used. Note that these derivatives hold for individuals choosing the bundle intended for them, as well as for mimickers.

5.2 Two Type Model

First, a two-type model is presented, with perfect correlation between w and a, where $w^2 > w^1$ and $a^2 > a^1$. The allocation is chosen to ensure that a type 2 individual does not opt for the bundle intended for type 1. The government offers bundles denoted as (Y, B_1, K, B_2) for both types. As in previous models, the government seeks to maximize the sum of utilities,

$$\begin{aligned} \max_{\{Y^i, B_1^i, K^i, B_2^i\}} \quad & \sum_i^2 n^i V^i, \\ \text{subject to} \quad & \sum_i^2 n^i (Y^i - B_1^i) \geq g_1 \quad (\lambda_1), \\ & \sum_i^2 n^i (K^i - B_2^i) \geq g_2 \quad (\lambda_2), \\ & V^2 \geq \hat{V}^{21} \quad (\gamma), \end{aligned} \tag{15}$$

where $\hat{V}^{21} = V(Y^1, B_1^1, K^1, B_2^1, w^2, a^2)$ denotes the indirect utility of a type 2 person mimicking a type 1 person, and $V^i = V(Y^i, B_1^i, K^i, B_2^i, w^i, a^i)$ as the indirect utility of a type i individual choosing the bundle intended for them. At the optimum, the above constraints will hold with equality. The corresponding Lagrange multipliers are in parentheses in (15).

All derivations are presented in Appendix B. The optimal allocation is described by

$$
\begin{aligned}
MRS_c^1 =& (1+k_s^1) - \frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2}\Big[(MRS_c^1 - M\hat{R}S_c^{21}) \\
&+ (1 - k_s^1/\hat{k}_s^{21})(M\hat{R}S_c^{21} - 1)\Big] < 1 + k_s^1,
\end{aligned} \tag{16}
$$

$$
MRS_Y^1 = 1 - \frac{\gamma\hat{u}_{21}'}{n^1\lambda_1}\Big[MRS_Y^1 - M\hat{R}S_Y^{21}\Big] \lesseqgtr 1. \tag{17}
$$

According to condition (16), capital income should be distorted downwards. If implemented with taxes, there would be a positive marginal tax on capital income. This result holds both when individuals only differ in terms of w and when they only differ in terms of a.[14] In a model where individuals only differ in terms of w, the only element that has been added to the standard two-period model that leads to the Atkinson-Stiglitz result is the possibility to exert investment effort. This means that the possibility to exert investment effort violates the Atkinson-Stiglitz result in an intertemporal setting, even if individuals have homogeneous investment ability.

The government wants to redistribute from the more skilled to the less skilled. More skilled individuals will, conditional on income, choose a different intertemporal allocation (they will front-load consumption) and they have a higher rate of return. Therefore, they are more willing to save at the margin. That is, the marginal propensity to save is higher for more skilled people. This means that the intertemporal allocation provides the government with information on the skill-level and should be used for taxation purposes.

As before, the distortion relaxes the incentive constraint. First, a mimicker and the type being mimicked have different intertemporal marginal rates of substitution. This reflects the first term in the bracket in (16). Second, a mimicker has a higher rate of return than the type being mimicked, i.e. $k_s^1 < \hat{k}_s^{21}$. This reflects the second term in the bracket in (16). Both effects imply that mimickers are more willing to increase savings. Therefore, distorting savings downwards makes mimicking less attractive.

The above result is similar to Proposition 1 in Gahvari & Micheletto (2016) and Proposition 2 in Gerritsen et al. (2020), where the optimal capital income tax is positive, as capital income provides the government with information on labour market productivity.

Condition (17) shows that the direction of the distortion on the labour-leisure decision of the type 1 individual is ambiguous. That is, the sign of the marginal tax rate on labour

[14] When individuals only differ in terms of w, the inequality in (16) is contingent on individuals exerting investment effort. If $E^i = 0\ \forall i$, there would be a corner solution, then capital income should not be distorted. The same does not apply when people differ only in terms of a. Then inequality in (16) also holds when $E^1 = \hat{E}^{21} = 0$. What matters for the distortion is that mimickers save less than the less able investor. From (14) it follows that mimickers save less even if $E^1 = \hat{E}^{21} = 0$, then $ds/da < 0$.

income can be positive or negative. This means that compared with the Mirrlees model, there is a mechanism that leads to a lower marginal tax rate on labour income. People with high labour market productivity front-load consumption. Therefore, they are less willing to increase their work effort than people with low productivity.

The reason that the distortion on labour supply is ambiguous is that there are two opposing forces, and it is ambiguous which will be stronger. As before, it matters whether the mimicker has a larger or lower MRS_Y than the low-skill individuals. If the mimicker has a lower (larger) MRS_Y than type 1 individual, there should be a downward (upward) distortion. First, mimickers save less than type 1 individuals, and therefore they need more compensation in terms of present consumption to earn one more unit of labour income. This calls for an upward distortion on labour supply. Second, mimickers have more leisure than the type mimicked (see comparative statics in Section 5.1). This calls for a downward distortion. In general, it is ambiguous which effect will be stronger, and therefore the direction of the inequality in (17) is ambiguous.

As in Section 4, marginal rates on labour income can be compared to those on capital income when individuals receive labour income in the second period, see Kristjánsson (2017). The results are unchanged, the government should impose larger marginal tax rates on the income type that is the predominant source of inequality.

Proposition 2 *If the government observes labour and capital income but not wealth, the optimal marginal tax rate on capital income is positive, while the sign on labour income is ambiguous.*

5.3 Four Type Model

Here, the general model with all four types of individuals is analysed. Due to the fact that $\frac{\partial V}{\partial w} > 0$ and $\frac{\partial V}{\partial a} > 0$, the government wants to redistribute from type 4 to types 1, 2 and 3, and from types 3 and 2 to type 1. This is shown by the direction of the arrows in both cases depicted in Figure 1. But the direction of redistribution between type 3 and 2 depends on the joint distribution of w and a. As the difference in $w^h - w^l$ becomes sufficiently large compared to $a^h - a^l$, case 1 applies, and vice versa.

In terms of potentially binding incentive constraints, there are two cases. In case 1(2), the government wants to perform redistribution from type 3(2) to type 2(3). Then the government needs to prevent type 3(2) to choose the bundle intended for type 2(3). This is shown in Figure 1 which shows all the possibly binding incentive constraints.

The general results from the four-type model are: In case 1(2), type 2(3) should be subject to a positive marginal tax rate on capital income, while the marginal tax rate on

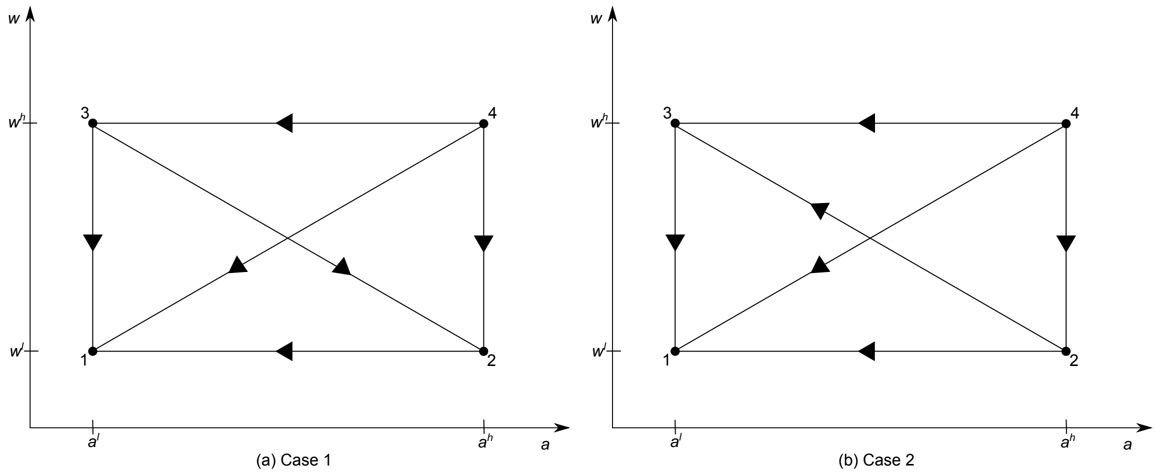

Figure 1: Potentially binding incentive constraints in case 1 and 2.

type 3(2) is ambiguous. Type 1 faces a positive marginal tax rate on capital income, see Appendix C.

When it is known which case applies, it is ambiguous how the mimicker will behave in comparison to the type being mimicked. The reason is that it cannot be shown in general whether e.g. type 3 mimicking type 2 will save more, and who has a higher rate of return. This is because there are two opposing forces. Type 3 mimickers have a higher w than type 2, which implies that they save less and have a higher k_s. However, they have a lower a than type 2, which implies that they save more and have a lower k_s. Which effect dominates depends on $w^h - w^l$ and $a^h - a^l$.

In what follows, I define $\Delta = w^h - w^l$ and allow it to vary while keeping $a^h - a^l$ fixed. When Δ is sufficiently small, then $V^2 > V^3$ and case 2 applies. Then $MRS_c^3 > M\hat{R}S_c^{23}$ and $k_s^3 < \hat{k}_s^{23}$.[15] Then, the optimal tax system distorts the intertemporal allocation of types 2 and 3 downwards, i.e. a positive marginal tax rate on capital income for type 2 and 3.

As Δ increases, there are three critical values:

- $\hat{\Delta}$: the value that corresponds to $V^2 = V^3$
- $\bar{\Delta}$: the value that corresponds to $s^3 = \hat{s}^{23}$
- $\widetilde{\Delta}$: the value that corresponds to $s^2 = \hat{s}^{32}$

When $s^2 = \hat{s}^{32}$, then $k_s^2 = \hat{k}_s^{32}$, and when $s^3 = \hat{s}^{23}$, then $k_s^3 = \hat{k}_s^{23}$.[16]

[15] Since $w^3 - w^2 > 0$ is small relative to $a^2 - a^3 > 0$, the higher a of type 2 compared to type 3 implies that $MRS_c^3 > M\hat{R}S_c^{23}$ and $k_s^3 < \hat{k}_s^{23}$, which calls for a downward distortion on type 3. Conversely, the higher w of type 3 compared to type 2 suggests the opposite effect. However, if Δ is sufficiently small, the former effect outweighs the latter.

[16] This relationship follows from the functional form assumed in (20).

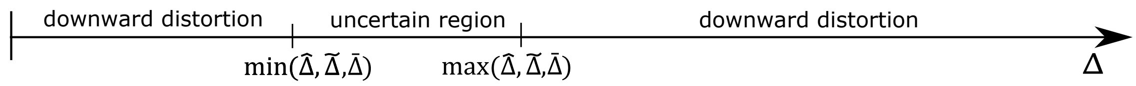

Figure 2: The direction of intertemporal distortions on type 2 and 3 when $\min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}) < \max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})$.

As Δ increases but remains below $\min(\hat{\Delta}, \bar{\Delta}, \widetilde{\Delta})$, we remain in case 2 and types 3 and 2 should face a downward distortion.[17] As Δ gets sufficiently high such that $\Delta > \max(\hat{\Delta}, \bar{\Delta}, \widetilde{\Delta})$, case 1 applies and $MRS_c^2 > M\hat{R}S_c^{32}$ and $k_s^2 < \hat{k}_s^{32}$. Then again, type 2 and 3 should face a downward distortion.

When Δ is in between these various critical values, i.e. $\Delta \in (\min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}), \max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}))$, then the direction of the distortion on either type 2 and 3 is unclear. When for example, $\hat{\Delta} < \min(\bar{\Delta}, \widetilde{\Delta})$, then $\gamma^{23} = 0$ and $\gamma^{32} > 0$, and $MRS_c^2 < M\hat{R}S_c^{32}$ and $k_s^2 > \hat{k}_s^{32}$. Then it is unclear whether the distortion on type 2 should be upward or downward. The smaller $\max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}) - \min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})$ is, the smaller the uncertain region depicted in figure 2 will be. Results are summarised in Proposition 3.

Proposition 3 *If the government observes capital and labour income, but not wealth, the optimal marginal tax rate on capital income is positive for types 1 to 3 if Δ is outside the uncertain region depicted in Figure 2. If Δ is within the uncertain region, the marginal tax rate on either type 2 or 3 is ambiguous.*

6 Financial advice

In this section, I expand the model to allow individuals to spend money to gain information, thereby achieving a higher rate of return. In practice, only a minority of individuals rely on financial advisors, and evidence suggests that this is complementary with financial knowledge. Those with higher levels of income, wealth, education, and financial literacy are the most likely to receive financial advice.[18] In the models by Delavande et al. (2008) and Lusardi et al. (2017), which are discussed in Section 3, financial advice and financial knowledge are complements. Piketty (2014) argue that this is the main reason returns are heterogeneous.

Financial advice is added to the model by modifying the capital income function to $k(E, m, s, a)$, where m denotes the expenditure on financial advice. Individuals can increase

[17] Then, $\gamma^{23} > 0$ and $\gamma^{32} = 0$, hence the downward distortions, see (21) in the Appendix C. Remember that type 1 always faces a downward distortion and type 4 always remains undistorted.

[18] See Ackerman et al. (2012), Collins (2012), and Finke (2013).

their rate of return in two ways: first, by exerting investment effort, which involves reducing leisure time; and second, by spending money in the first period, at the cost of reducing consumption during that period. The capital income function, which is increasing in m, has the following properties, other properties remain as described in Section 3,

$$k_m, k_{sm}, k_{am}, k_{Em} > 0,\ k_{mm} \leq 0.$$

An increase in m leads to a higher rate of return, as indicated by $k_{sm} > 0$. Seeking financial advice leads to an increase in capital income, albeit at a diminishing rate. Financial advice also increases the return to savings and the return due to investment ability.

Individuals pay the financial advisor in the first period, which increases capital income in the second period. The individual budget constraints for periods 1 and 2 are now, respectively,

$$\begin{aligned} c_1 &= Y - s - m - t, \\ c_2 &= s + k(E, m, s, a) - T. \end{aligned}$$

Optimal allocations are derived in a two-type model with perfect correlation between w and a, as in Sections 4 and 5.2. The government observes Y, s, and K, but not m.[19] All the comparative statics, the government's problem, and the necessary conditions are presented in Appendix D.

The crucial question for optimal taxation is whether a mimicker spends more or less on financial advisors compared to the type being mimicked, and whether the mimicker has more leisure than the type being mimicked. From Appendix D, it can be seen that mimickers spend less on financial advice, and therefore enjoy more first-period consumption, and therefore their k_m is higher, and they also enjoy more leisure.

The optimal intertemporal allocation is characterized by the following conditions:

$$MRS_c^1 = 1 + k_s^1 - \frac{\gamma\psi_1'}{n^1\lambda_2}\left[MRS_c^1 - M\hat{R}S_c^{21}\right] < 1 + k_s^1, \tag{18a}$$

$$MRS_c^1 = k_m^1 - \frac{\gamma\psi_1'}{n^1\lambda_2}\left[MRS_c^1 - M\hat{R}S_c^{21}\frac{k_m^1}{\hat{k}_m^{21}}\right] < k_m^1. \tag{18b}$$

The inequalities indicate that the intertemporal allocation should be distorted downwards, as $MRS_c^i = 1 + k_s^i = k_m^i$ in the first-best scenario. As before, the government aims to

[19] The main results also hold when m is tax-deductible, i.e. when the government observes $k(E, m, s, a) - m$. If m were observable, optimal allocations were such that the intertemporal allocation should be left undistorted, similar to Section 4.

redistribute resources from the skilled to the less skilled. Individuals with higher innate abilities, conditional on income, have higher first-period consumption and, therefore, are more inclined to save. This implies that the intertemporal allocation provides the government with information on innate ability, which should be utilised for tax purposes. Moreover, those with greater innate ability achieve a higher return on spending on investment advisors, as they spend less on financial advice. This calls for a higher distortion on the financial advisory decision compared to the savings decision, see Appendix D.

To analyse how the distortions characterised in (18) can be achieved by a tax system, consider a tax function as in (7):

$$c_2 = s + k(E, s, m, a) - T(s + k(E, s, m, a), k(E, s, m, a)). \tag{19}$$

An increase in financial advice equally affects the wealth tax base and the capital income tax base, increasing both tax bases by k_m. An increase in savings, however, does not equally affect the tax bases. The wealth tax base increases by $1 + k_s$, whereas the capital income tax base increases by k_s.

Since the capital income tax hits financial advice relatively harder than a wealth tax does, the capital income tax may play a more important role than the wealth tax. In general, though, the sign of the marginal tax rates cannot be determined. However, if the wealth tax only consists of savings, then the capital income tax rate is unambiguously positive, and the marginal wealth tax is positive if the marginal capital income tax rate were sufficiently large, see Appendix D.

Proposition 4 *When individuals can spend money on financial advisors and the government observes savings, as well as capital and labour income, the optimal tax system is such that the intertemporal allocation is distorted downwards.*

7 Conclusion

I have analysed optimal nonlinear taxation of labour income and capital in a two-period model where individuals can exert investment effort as well as supply labour. Individuals differ in labour market productivity and investment ability. In the Atkinson-Stiglitz model, capital income does not reveal information about underlying skills. Individuals who have high capital income do so because they saved a lot from high earnings. In my model, high capital income is due to savings, investment effort, investment ability, and financial advice. More skilled individuals are, conditional on income, more willing at the margin to save, exert investment effort, and spend money on financial advice. Therefore, capital income should be distorted, as it provides information on the underlying skill level.

A distinct feature of the model is that a capital income tax and a wealth tax have different effects. In a model with perfect capital markets, it makes no difference whether the government taxes capital income or wealth. For example, a 30% tax on capital income with a return of 5% is equivalent to a 1.5% tax on wealth. This is not the case in my model. In the baseline model, the marginal tax rate on capital income is positive, while the marginal tax rate on wealth is negative. As returns reflect the underlying skill-level, they provide a rationale for taxing capital income, which may then be more suitable than a wealth tax. In the model by Guvenen et al. (2023), another mechanism is at play. Optimal capital income taxes are negative while the wealth tax is positive. This encourages reallocation from less productive to more productive uses of capital. In my model, nonlinear taxation together with decreasing returns to scale ensure efficient allocation of savings.

The key to my result that capital income should be taxed lies in the market imperfections embedded in the model. When the government does not observe savings and nonlinear taxation does not ensure efficient allocation, individuals would benefit from interpersonal lending, which would equalise returns. In Kristjánsson (2017), I consider an extension where there is a domestic credit market. This somewhat deviates from the initial motivation. In a model with this extension, insights from Sections 4–6 remain valid, but there will be additional effects on the domestic interest rate that need to be taken into account.

References

Ackerman, R. A., Fries, G., & Windle, R. A. (2012). Changes in us family finances from 2007 to 2010: Evidence from the survey of consumer finances. *Federal Reserve Bulletin*, *100*(4), 1–80.

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

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

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

Banks, J., & Diamond, P. (2010). The base for direct taxation. In J. A. Mirrlees & S. Adam (Eds.), *Dimensions of tax design: The mirrlees review* (p. 548-648). Oxford: Oxford University Press.

Bastani, S., Karlsson, K., & Waldenström, D. (2023). *The different returns to cognitive ability in the labor and capital markets* (Tech. Rep.). IFAU-Institute for Evaluation of Labour Market and Education Policy.

Bastani, S., & Waldenström, D. (2020). How should capital be taxed? *Journal of Economic Surveys*, *34*(4), 812–846.

Bastani, S., & Waldenström, D. (2023). Taxing the wealthy: the choice between wealth and capital income taxation. *Oxford Review of Economic Policy*, *39*(3), 604–616.

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

Boadway, R., & Spiritus, K. (2021). Optimal taxation of normal and excess returns to risky assets. *Tinbergen Institute Discussion Papers 21-025/VI, Tinbergen Institute*.

Bovenberg, A. L., & Jacobs, B. (2005). Redistribution and education subsidies are siamese twins. *Journal of Public Economics*, *89*(11-12).

Campbell, J. Y. (2006). Household finance. *The journal of finance*, *61*(4), 1553–1604.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica: Journal of the Econometric Society*, 607–622.

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

Christiansen, V., & Tuomala, M. (2008). On taxing capital income with income shifting. *International Tax and Public Finance*, *15*, 527–545.

Collins, J. M. (2012). Financial advice: A substitute for financial literacy? *Financial services review*, *21*(4), 307.

Cremer, H., Lozachmeur, J.-M., & Pestieau, P. (2004). Social security, retirement age and optimal income taxation. *Journal of Public Economics*, *88*(11), 2259–2281.

Delavande, A., Rohwedder, S., & Willis, R. J. (2008). Preparation for retirement, financial literacy and cognitive resources. *Michigan Retirement Research Center Research Paper*(2008-190).

De Nardi, M., & Fella, G. (2017). Saving and wealth inequality. *Review of Economic Dynamics*, *26*, 280–300.

Deuflhard, F., Georgarakos, D., & Inderst, R. (2019). Financial literacy and savings account returns. *Journal of the European Economic Association*, *17*(1), 131–164.

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

Finke, M. (2013). Financial advice: Does it make a difference? In O. S. Mitchell & K. Smetters (Eds.), *The market for retirement financial advice.* Oxford, UK: Oxford University Press.

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

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

Gaillard, A., & Wangner, P. (2021). Wealth, returns, and taxation: A tale of two dependencies. *Available at SSRN*, *3966130*.

Gerritsen, A., Jacobs, B., Rusu, A. V., & Spiritus, K. (2020). Optimal taxation of capital income with heterogeneous rates of return.

Guiso, L., & Jappelli, T. (2008). Financial literacy and portfolio choice. *EUI Working o Papers, ECO2008/31*.

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S., & Chen, D. (2023). Use it or lose it: Efficiency and redistributional effects of wealth taxation. *The Quarterly Journal of Economics*, *138*(2), 835–894.

Hastings, J. S., Madrian, B. C., & Skimmyhorn, W. L. (2013). Financial literacy, financial education, and economic outcomes. *Annu. Rev. Econ.*, *5*(1), 347–373.

Jacobs, B. (2013). From optimal tax theory to applied tax policy. *FinanzArchiv/Public Finance Analysis*, 338–389.

Jacobs, B., & Bovenberg, A. L. (2010). Human capital and optimal positive taxation of capital income. *International Tax and Public Finance*, *17*, 451–478.

Jacquet, L., & Lehmann, E. (2023). Optimal tax problems with multidimensional heterogeneity: a mechanism design approach. *Social Choice and Welfare*, 1–30.

Jappelli, T., & Padula, M. (2017). Consumption growth, the interest rate, and financial sophistication. *Journal of Pension Economics & Finance*, *16*(3), 348–370.

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

Kristjánsson, A. S. (2017). Optimal taxation with endogenous return to capital. *Working Paper 06/2016, Oslo University*.

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

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *American Economic Journal: Journal of Economic Literature*, *52*(1), 5–44.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, *38*(2), 175–208.

Piketty, T. (2014). *Capital in the twenty-first century.* Harvard University Press.

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

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

Schulz, K. (2021). Redistribution of return inequality.

Slemrod, J., & Gillitzer, C. (2014). *Tax systems.* MIT Press Cambridge, MA.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, *17*(2), 181–211.
Stiglitz, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of public economics*, *17*(2), 213–240.
Stiglitz, J. E. (1985). Inequality and capital taxation. *IMSSS Technical Report No. 457*.
Stiglitz, J. E. (1987). Pareto efficient and optimal taxation and the new new welfare economics. In *Handbook of public economics* (Vol. 2, pp. 991–1042). Elsevier.
Stolper, O. A., & Walter, A. (2017). Financial literacy, financial advice, and financial behavior. *Journal of business economics*, *87*, 581–643.
Straub, L., & Werning, I. (2020). Positive long-run capital taxation: Chamley-judd revisited. *American Economic Review*, *110*(1), 86–119.
Xavier, I. (2021). Wealth inequality in the us: the role of heterogeneous returns. *Available at SSRN 3915439*.
Yitzhaki, S. (1987). The relation between return and income. *The Quarterly Journal of Economics*, *102*(1), 77–95.
Zanoutene, E. (2023). Scale-dependent and risky returns to savings: Consequences for optimal capital taxation. *Journal of Public Economic Theory*, *25*(3), 532–569.

Appendix

A Calculations for Section 4

Necessary Conditions The Lagrangian for the government's problem is

$$\mathcal{L} = \sum_i n^i U^i + \gamma \left[U^2 - \hat{U}^{21} \right] + \lambda_1 \left[\sum_i n^i (Y^i - B_1^i) - g_1 \right] + \lambda_2 \left[\sum_i n^i (k(E^i, a^i, s^i) - B_2^i) - g_2 \right].$$

From the condition $k(E^1, a^1, s^1) - k(\hat{E}^{21}, a^2, s^1) = 0$, it follows that

$$\frac{d\hat{E}^{21}}{dE^1} = \frac{k_E^1}{\hat{k}_E^{21}}, \quad \text{and} \quad \frac{d\hat{E}^{21}}{ds^1} = \frac{k_s^1 - \hat{k}_s^{21}}{\hat{k}_E^{21}}.$$

If the following form of weak separability is satisfied,

$$k(E, s, a) = f(E)g(s)h(a), \tag{20}$$

then $k_s^1 = \hat{k}_s^{21}$. When individuals only differ in terms of w, then, $\hat{E}^{21} = E^1$ and $k_s^1(E^1, s^1, a) = \hat{k}_s^{21}(\hat{E}^{21}, s^1, a)$ follows irrespective of $k(\cdot)$. When individuals only differ in terms of a, then $\frac{dk_s}{da} = k_{sE}\frac{dE}{da} + k_{sa} = -k_{sE}\frac{k_a}{k_E} + k_{sa} = 0$. The form (20) of weak separability will be assumed to be the case from now on.

The necessary conditions for type 1 are

$$\begin{aligned}
(Y^1) \qquad & -n^1 v_1'/w^1 + \gamma \hat{v}_{21}'/w^2 + \lambda_1 n^1 = 0, \\
(E^1) \qquad & -n^1 v_1' + \gamma \hat{v}_{21}' k_E^1/\hat{k}_E^{21} + \lambda_2 n^1 k_E^1 = 0, \\
(s^1) \qquad & (n^1 - \gamma)(\psi_1' - u_1') + \lambda_2 n^1 k_s^1 = 0, \\
(B_1^1) \qquad & (n^1 - \gamma)u_1' - \lambda_1 n^1 = 0, \\
(B_2^1) \qquad & (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0,
\end{aligned}$$

To get (6a), I solve for v_1'/k_E^1 and ψ_1' from the necessary conditions for E^1 and B_2^1, respectively, and divide. Next, I multiply both sides with $(\gamma\psi_1' + \lambda_2 n^1)/\gamma\psi_1'$ and after simple manipulations, I get (6a). Equation (6c) is found by very similar algebraic steps.

As discussed in the beginning of section 4, mimickers have more leisure than type 1 individual, hence $v_1' > \hat{v}_{21}'$. Since $a^2 > a^1$, $k_E^1(E^1, a^1, s^1) < \hat{k}_E^{21}(\hat{E}^{21}, a^2, s^1)$, since $k_{EE} \leq 0$ and $k_{Ea} > 0$. Hence, $MRS_K^1 = \frac{v_1'}{\psi_1' k_E^1} > M\hat{R}S_K^{21} = \frac{\hat{v}_{21}'}{\hat{\psi}_{21}' \hat{k}_E^{21}}$, and $MRS_Y^1 = \frac{v_1'}{u_1' w^1} > M\hat{R}S_Y^{21} = \frac{\hat{v}_{21}'}{\hat{u}_{21}' w^2}$.

Difference in Distortion between Y and K Since $MRS_K = MRS_Y \frac{w(1+k_s)}{k_E}$, then $\text{sign}(t' - T_K) = \text{sign}(w(1 + k_s) - k_E)$. This depends on $w^2 - w^1$ and $a^2 - a^1$. Using the necessary condition for Y^1 and E^1, I can write the following expression:

$$\lambda_2 n^1 [w^1(1 + k_s^1) - k_E^1] = \gamma \hat{v}_{21}' \left[\frac{k_E^1}{\hat{k}_E^{21}} - \frac{w^1}{w^2} \right].$$

Due to the necessary conditions from footnote 13: $MRS_Y = 1 - t'$ and $MRS_K = 1 - T_K$, it follows that,

$$\begin{aligned}
&(w^2 > w^2, a^2 = a^1): \qquad && w^1(1 + k_s^1) > k_E^1 \;\Rightarrow\; MRS_K^1 < MRS_Y^1 \;\Rightarrow\; t' > T_K, \\
&(w^2 = w^2, a^2 > a^1): \qquad && w^1(1 + k_s^1) < k_E^1 \;\Rightarrow\; MRS_K^1 > MRS_Y^1 \;\Rightarrow\; t' < T_K.
\end{aligned}$$

B Calculations for Section 5.2

Condition for $s_{Ea} = 0$ Given the properties of $k(E, s, a)$ in (20), $s_{Ea} = (k_E k_{sa} - k_s k_{Ea})/k_s^2 = 0$.

Necessary Conditions The Lagrangian for (15) is

$$\mathcal{L} = \sum_i n^i V^i + \gamma \left[V^2 - \hat{V}^{21} \right] + \lambda_1 \left[\sum_i n^i (Y^i - B_1^i) - g_1 \right] + \lambda_2 \left[\sum_i n^i (K^i - B_2^i) - g_2 \right].$$

The necessary conditions for type 1 are

$$\begin{aligned}
&(Y^1) \qquad && - n^1 v_1'/w^1 + \gamma \hat{v}_{21}'/w^2 + \lambda_1 n^1 = 0, \\
&(K^1) \qquad && - n^1 u_1'/k_s^1 + n^1 \psi_1'/k_s^1 + \gamma \hat{u}_{21}'/\hat{k}_s^{21} - \gamma \hat{\psi}_{21}'/\hat{k}_s^{21} + \lambda_2 n^1 = 0, \\
& && - n^1 v_1'/k_E^1 + \gamma \hat{v}_{21}'/\hat{k}_E^{21} + \lambda_2 n^1 = 0, \\
&(B_1^1) \qquad && n^1 u_1' - \gamma \hat{u}_{21}' - \lambda_1 n^1 = 0, \\
&(B_2^1) \qquad && n^1 \psi_1' - \gamma \hat{\psi}_{21}' - \lambda_2 n^1 = 0.
\end{aligned}$$

Condition (16) I solve for u_1' and ψ_1' from the necessary conditions for K^1 and B_2^1, respectively, and divide them together. Next, I multiply both sides by $(\gamma \hat{\psi}_{21}' + \lambda_2 n^1)/\lambda_2 n^1$ and rearrange,

$$\frac{u_1'}{\psi_1'} = \frac{\psi_1'}{\lambda_2} + k_s^1 - \frac{u_1'}{\psi_1'} \frac{\gamma \hat{\psi}_{21}'}{n^1 \lambda_2} + \frac{\gamma \hat{\psi}_{21}'}{n^1 \lambda_2} \frac{k_s^1}{\hat{k}_s^{21}} \left(\frac{\hat{u}_{21}'}{\hat{\psi}_{21}'} - 1 \right).$$

Noting $\psi_1'/\lambda_2 = \gamma \hat{\psi}_{21}'/\lambda_2 n^1 + 1$ from the necessary condition for B_2^1 and rearranging will give (16).

In order to show the inequality in (16), first $\hat{k}_s^{21} > k_s^1$ needs to be established. This is done by differentiating $k_s(E(Y, B_1, K, B_2, w, a), s(E(Y, B_1, K, B_2, w, a), a, K), a)$ w.r.t. a and w. Here, $E(Y, B_1, K, B_2, w, a)$ is the individual's optimal choice, which is analyses by differentiating (10) and $s(E, a, K)$, which follows from the constraint (9),

$$\begin{aligned}
\frac{dk_s}{dw} &= k_{sE} \frac{dE}{dw} + k_{ss} s_E \frac{dE}{dw} > 0, \\
\frac{dk_s}{da} &= k_{sE} \frac{dE}{da} + k_{ss} \left[s_E \frac{dE}{da} + s_a \right] + k_{sa} \\
&= \frac{-k_{sE} U_{Ea} - k_{ss} s_E U_{Ea} + k_{ss} s_a U_{EE} + k_{sa} U_{EE}}{U_{EE}}
\end{aligned}$$

$$= \frac{\left[k_{ss}s_a + k_{sa}\right]\left[v'' - (u' + \psi')s_{EE}\right]}{U_{EE}} + \frac{(u'' + \psi'')s_E}{U_{EE}}\left[\frac{1}{k_s}(k_{sE}k_a - k_{sa}k_E)\right.$$
$$\left.+k_{ss}(s_E s_a - s_E s_a)\right] = \frac{\left[k_{ss}s_a + k_{sa}\right]\left[v'' - (u' + \psi')s_{EE}\right]}{U_{EE}} > 0,$$

where $\frac{dE}{dw} > 0$ follows from (10) and $s_E < 0$ from (9), and $k_{sE}k_a - k_{sa}k_E = 0$ is due to (20). Since k_s is increasing in w and a, mimickers will have a higher k_s, hence $\hat{k}_s^{21} > k_s^1$.

Next, the signs of $MRS_c^1 - \hat{MRS}_c^{21}$ need to be established. Since mimickers save less than the type being mimicked ($s^1 > \hat{s}^{21}$) it follows that $u_1' > \hat{u}_{21}'$ and $\psi_1' < \hat{\psi}_{21}'$, hence $\frac{u_1'}{\psi_1'} = MRS_c^1 > \frac{\hat{u}_{21}'}{\hat{\psi}_{21}'} = \hat{MRS}_c^{21}$.

Finally, the sign of $\hat{MRS}_c^{21} - 1$ need to be established. From the individual's necessary condition (10), it follows that $u' - \psi' = v'E_s > 0$, hence $\frac{u'}{\psi'} > 1$, for mimickers and non-mimickers, hence $\hat{MRS}_c^{21} - 1 > 0$.

The above shows that the bracket in (16) is positive both when individuals only differ in terms of w and when they only differ in terms of a. Since $\frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2} > 0$, then $MRS_c^1 < 1 + k_s^1$.

Condition (17) I solve for v_1' and u_1' from the necessary conditions for Y^1 and B_1^1, respectively, and divide them together. Next, I multiply both sides with $(\gamma\hat{u}_{21}' + \lambda_1 n^1)/\lambda_1 n^1$ and after some manipulation I get (17).

In order to derive the inequality in (17) for both heterogeneous a and heterogeneous w. As $s^1 > \hat{s}^{21}$ and $E^1 + L^1 > \hat{E}^{21} + \hat{L}^{21}$, hence $u_1' > \hat{u}_{21}'$ and $v_1' > \hat{v}_{21}'$. The sign of $MRS_Y^1 - \hat{MRS}_Y^{21}$ is ambiguous, and therefore the direction of the inequality in (17) is ambiguous.

C Calculations for Section 5.3

The government's problem is identical to the two-type case, except that there are now four types and four additional incentive constraints, with at most three of them binding. The Lagrangian for the problem is

$$\begin{aligned}
\mathcal{L} = &\sum_i^4 n^i V^i + \gamma^{21}\left[V^2 - \hat{V}^{21}\right] + \gamma^{31}\left[V^3 - \hat{V}^{31}\right] \\
&+ \left(\gamma^{32}\left[V^3 - \hat{V}^{32}\right] + \gamma^{23}\left[V^2 - \hat{V}^{23}\right]\right) + \gamma^{42}\left[V^4 - \hat{V}^{42}\right] + \gamma^{43}\left[V^4 - \hat{V}^{43}\right] \\
&+ \lambda_1\left[\sum_i^4 n^i(Y^i - B_1^i) - g_1\right] + \lambda_2\left[\sum_i^4 n^i(K^i - B_2^i) - g_2\right],
\end{aligned}$$

where one of the two incentive constraints in the parenthesis is not binding, i.e. either $\gamma^{32} = 0$ or $\gamma^{23} = 0$, or both. In analysing the solution to this problem I will not look at the intratemporal allocation since there are the same forces at play as in the two type models and it also not possible to derive the signs of wedges (i.e. whether labour should be taxed or subsidised at the margin).

To simplify, I will assume that the incentive constraint on type 4 mimicking type 1 is slack (this will not affect the main qualitative results). The necessary conditions are

$$
\begin{aligned}
&(Y^1) && -n^1 v_1'/w^l + \gamma^{21}\hat{v}_{21}'/w^l + \gamma^{31}\hat{v}_{31}'/w^h + \lambda_1 n^1 = 0,\\
&(Y^2) && -(n^2+\gamma^{21}+\gamma^{23})v_2/w^l - \gamma^{32}\hat{v}_{32}'/w^h - \gamma^{42}\hat{v}_{42}'/w^h + \lambda_1 n^2 = 0,\\
&(Y^3) && -(n^3+\gamma^{31}+\gamma^{32})v_3'/w^h - \gamma^{23}\hat{v}_{23}'/w^l - \gamma^{43}\hat{v}_{43}'/w^h + \lambda_1 n^3 = 0,\\
&(K^1) && -n^1(u_1'-\psi_1')/k_s^1 + \gamma^{21}(\hat{u}_{21}'-\hat{\psi}_{21}')/\hat{k}_s^{21} + \gamma^{31}(\hat{u}_{31}'-\hat{\psi}_{31}')/\hat{k}_s^{31} + \lambda_2 n^1 = 0,\\
&(K^2) && -(n^2+\gamma^{21}+\gamma^{23})(u_2'-\psi_2')/k_s^2 + \gamma^{32}(\hat{u}_{32}'-\hat{\psi}_{32}')/\hat{k}_s^{32}\\
& && +\gamma^{42}(\hat{u}_{42}'-\hat{\psi}_{42}')/\hat{k}_s^{42} + \lambda_2 n^2 = 0,\\
&(K^3) && -(n^3+\gamma^{31}+\gamma^{32})(u_3'-\psi_3')/k_s^3 + \gamma^{23}(\hat{u}_{23}'-\hat{\psi}_{23}')/\hat{k}_s^{23}\\
& && +\gamma^{43}(\hat{u}_{43}'-\hat{\psi}_{43}')/\hat{k}_s^{43} + \lambda_2 n^3 = 0,\\
&(B_1^1) && n^1 u_1' - \gamma^{21}\hat{u}_{21}' - \gamma^{31}\hat{u}_{31}' - \lambda_1 n^1 = 0,\\
&(B_1^2) && (n^2+\gamma^{21}+\gamma^{23})u_2' - \gamma^{32}\hat{u}_{32}' - \gamma^{42}\hat{u}_{42}' - \lambda_1 n^2 = 0,\\
&(B_1^3) && (n^3+\gamma^{31}+\gamma^{32})u_3' - \gamma^{23}\hat{u}_{23}' - \gamma^{43}\hat{u}_{43}' - \lambda_1 n^3 = 0,\\
&(B_2^1) && n^1 \psi_1' - \gamma^{21}\hat{\psi}_{21}' - \gamma^{31}\hat{\psi}_{31}' - \lambda_2 n^1 = 0,\\
&(B_2^2) && (n^2+\gamma^{21}+\gamma^{23})\psi_2' - \gamma^{32}\hat{\psi}_{32}' - \gamma^{42}\hat{\psi}_{42}' - \lambda_2 n^2 = 0,\\
&(B_2^3) && (n^3+\gamma^{31}+\gamma^{32})\psi_3' - \gamma^{23}\hat{\psi}_{23}' - \gamma^{43}\hat{\psi}_{43}' - \lambda_2 n^3 = 0.
\end{aligned}
$$

The necessary conditions as in Appendix A and B lead to the following optimal intertemporal allocations

$$
\begin{aligned}
MRS_c^1 =& (1+k_s^1) - \frac{\gamma^{21}\hat{\psi}_{21}'}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{21}) + (1-k_s^1/\hat{k}_s^{21})(M\hat{R}S_c^{21}-1)\right]\\
& - \frac{\gamma^{31}\hat{\psi}_{31}'}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{31}) - (1-k_s^1/\hat{k}_s^{31})(M\hat{R}S_c^{31}-1)\right] < 1+k_s^1, \qquad (21a)\\
MRS_c^2 =& (1+k_s^2) - \frac{\gamma^{32}\hat{\psi}_{32}'}{n^2\lambda_2}\left[(MRS_c^2 - M\hat{R}S_c^{32}) - (1-k_s^2/\hat{k}_s^{32})(M\hat{R}S_c^{32}-1)\right]\\
& - \frac{\gamma^{42}\hat{\psi}_{42}'}{n^2\lambda_2}\left[(MRS_c^2 - M\hat{R}S_c^{42}) - (1-k_s^2/\hat{k}_s^{42})(M\hat{R}S_c^{42}-1)\right]\\
& < 1+k_s^1,\ \text{if } \gamma^{32}=0,\ \text{or if } \gamma^{32}>0,\ MRS_c^2 > M\hat{R}S_c^{32},\ \text{and } k_s^2 < \hat{k}_s^{32}, \qquad (21b)\\
MRS_c^3 =& (1+k_s^3) - \frac{\gamma^{23}\hat{\psi}_{23}'}{n^3\lambda_2}\left[(MRS_c^3 - M\hat{R}S_c^{23}) - (1-k_s^3/\hat{k}_s^{23})(M\hat{R}S_c^{23}-1)\right]\\
& - \frac{\gamma^{43}\hat{\psi}_{43}'}{n^3\lambda_2}\left[(MRS_c^3 - M\hat{R}S_c^{43}) - (1-k_s^3/\hat{k}_s^{43})(M\hat{R}S_c^{43}-1)\right]\\
& < 1+k_s^1 \text{ if } \gamma^{23}=0,\ \text{or if } \gamma^{23}>0,\ MRS_c^3 > M\hat{R}S_c^{23},\ \text{and } k_s^3 < \hat{k}_s^{23}. \qquad (21c)
\end{aligned}
$$

The inequalities can be proven analogously to Appendix A and no need to repeat the calculations here. This means that at least two types will be distorted downwards, assuming a separating equilibrium. Either $\gamma^{32} = 0$ or $\gamma^{23} = 0$ depending on the joint distribution of w and a as well as the bundles that are offered.

D Calculations for Section 6

Comparative Statics The government observes Y, s, and K, and offers bundles in terms of these variables for both types of individuals. I follow a similar procedure as in Section 5.2. Individuals choose E and m, facing the constraint $k(E, s, m, a) = K$. This implicitly defines $E(K, s, m)$. Partial derivatives are derived by implicitly differentiating $k(E, s, m, a) - K = 0$. The individual's problem and necessary condition are, respectively,

$$\begin{aligned}
&\max_{\{m\}} U = u(B_1 - s - m) + \psi(B_2 + s) + v(1 - Y/w - E(m, s, K, a)),\\
&U_m = -u'(B_1 - s - m) - v'(1 - Y/w - E(m, s, K, a))E_m(m, s, K, a) = 0,
\end{aligned}$$

where $E_m = -k_m/k_E < 0$. As before: $B_1 = Y - t$ and $B_2 = K - T$.

Implicitly differentiating the condition $U_m = 0$ gives the following comparative static results for w:

$$\begin{aligned}
\frac{dm}{dw} =& \frac{-U_{mw}}{U_{mm}} = \frac{-v''Yw^{-2}E_m}{u'' - v'E_{mm} + v''E_m^2} < 0,\\
\frac{dE}{dw} =& E_m\frac{dm}{dw} = \frac{-v''E_m^2}{u'' - v'E_{mm} + v''E_m^2}\frac{Y}{w^2} > 0 \longrightarrow \frac{dE}{dw} < -\frac{dL}{dw} = \frac{Y}{w^2},
\end{aligned}$$

where $E_{mm} = [(k_m k_{Em} - k_E k_{EE}]/k_E^2 > 0$. An increase in w leads to a decrease in m and an increase in E. The mechanical effect of mimicking is that L is lower than that for the type being mimicked. Due to the concavity in the utility function, a mimicker wants to balance the utility gain of more leisure by increasing c_1 and increasing leisure by less than $\frac{dL}{dw}$. This means that mimickers differing in terms of w enjoy more leisure and more c_1.

Performing the same comparative statics for an increase in a gives:

$$\begin{aligned}
\frac{dm}{da} =& \frac{v'E_{ma} - v''E_m E_a}{U_{mm}} < 0,\\
\frac{dE}{da} =& E_m\frac{dm}{da} + E_a = \frac{u''E_a + v'[E_{ma}E_m - E_{mm}E_a]}{U_{mm}} = \frac{E_a u''}{U_{mm}} - \frac{v'k_{mm}k_a}{U_{mm}k_E^2} < 0
\end{aligned}$$

where $E_a = -k_a/k_E < 0$ and $E_{ma} = k_m k_{Ea}/k_E^2 > 0$. An increase in a leads to a decrease in m and E. Mimickers differing in terms of a enjoy more leisure and more c_1.

The derivatives of the indirect utility function are

$$\frac{\partial V}{\partial K} = -\frac{v'}{k_E} = -\frac{u'}{k_m}, \qquad \frac{\partial V}{\partial s} = -u' + \psi' + v'\frac{k_s}{k_E} = -u'\left(1 - \frac{k_s}{k_m}\right) + \psi', \qquad \frac{\partial V}{\partial B_2} = \psi'.$$

Government's Necessary Conditions The Lagrangian for the government's problem is

$$\mathcal{L} = \sum_i n^i V^i + \gamma \left[V^2 - \hat{V}^{21}\right] + \lambda_1 \left[\sum_i n^i (Y^i - B_1^i) - g_1\right] + \lambda_2 \left[\sum_i n^i (K^i - B_2^i) - g_2\right].$$

The necessary conditions for type 1 are

$$(K^1) \qquad -n^1 u_1'/k_m^1 + \gamma \hat{u}_{21}'/\hat{k}_m^{21} + \lambda_2 n^1 = 0, \tag{22a}$$

$$(K^1) \qquad -n^1 v_1'/k_E^1 + \gamma \hat{v}_{21}'/\hat{k}_E^{21} + \lambda_2 n^1 = 0 \tag{22a$'$}$$

$$(s^1) \qquad -n^1 u_1' \left(1 - k_s^1/k_m^1\right) + n^1 \psi_1' + \gamma \hat{u}_{21}' \left(1 - \hat{k}_s^{21}/\hat{k}_m^{21}\right) - \gamma \psi_1' = 0, \tag{22b}$$

$$(B_2^1) \qquad (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0. \tag{22c}$$

Solving for $n_i u_i'$ from (22a) and for $n_i \psi_i'$ in (22c), and performing similar derivations as in Appendix B yields condition (18b). Solving for $n^1 u_1'$ from (22b) and plugging $n^1 u_1'/k_m^1$ into the conditions yields

$$n^1 u_1' = \lambda_2 n^1 k_s^1 + \frac{\gamma \hat{u}_{21}'}{\hat{k}_m^{21}} k_s^1 + \psi'(n^1 - \gamma) + \gamma \hat{u}_{21}' \left(1 - \frac{k_s^1}{k_m^1}\right).$$

Next, using $(n^1 - \gamma)\psi_1' = \lambda_2 n^1$ from (22c), yields

$$n^1 u_1' = \lambda_2 n^1 (1 + k_s^1) + \gamma \hat{u}_{21}' + \frac{\gamma \hat{u}_{21}'}{\hat{k}_m^{21}} (k_s^1 - \hat{k}_s^{21}).$$

Solving for $n^1 \psi_1'$ from (22c) and dividing with the above condition, and performing similar derivations as in Appendix B gives condition (18a) if the capital income function is weakly separable, similar to (20),

$$k(E, s, m, a) = f(E) g(s) j(m) h(a), \tag{23}$$

which will be assumed to be the case. If (23) holds, then

$$\begin{aligned}
\frac{dk_s}{dw} =& k_{sE}\frac{dE}{dw} + k_{sm}\frac{dm}{dw} = \frac{dm}{dw}\left[k_{sm} - \frac{k_{sE}k_m}{k_E}\right] = 0, \\
\frac{dk_s}{da} =& k_{sE}\frac{dE}{da} + k_{sm}\frac{dm}{da} + k_{sa} = \frac{v'' E_m}{U_{mm} k_E}\left[k_{sm}k_a - k_{sa}k_m\right] + \frac{u''}{U_{mm}}\left[k_{sa} - \frac{k_{sE}k_a}{k_E}\right] \\
&+ \frac{v'}{U_{mm}}\left[E_{ma}\left(k_{sm} - \frac{k_{sE}k_m}{k_E}\right) - E_{mm}\left(\frac{k_{sE}k_a}{k_E} - k_{sa}\right)\right] = 0,
\end{aligned}$$

where all brackets above are zero, which follows from the weakly separability in (23). This shows that $\hat{k}_s^{21} - k_s^1 = 0$.

Finally, it needs to be established whether $\frac{k_m^1}{\hat{k}_m^{21}} \lesseqgtr 1$, which can be done by performing the following derivatives:

$$\frac{dk_m}{dw} = k_{mE}\frac{dE}{dw} + k_{mm}\frac{dm}{dw} > 0,$$
$$\frac{dk_m}{da} = k_{mE}\frac{dE}{da} + k_{mm}\frac{dm}{da} + k_{ma} = \frac{v''(E_m k_{ma-} k_{mm} E_a) - v' k_{ma} E_{mm}}{U_{mm}} > 0.$$

This shows that $k_m^1 < \hat{k}_m^{21}$, both if individuals differ in terms of w and in terms of a.

An exogenous increase in w leads to a reduction in m and an increase in E. Both will increase k_m, since $k_{EE} < 0$ and $k_{mE} > 0$. An exogenous increase in a will mechanically increase k_m, since $k_{ma} > 0$. An increase in a leads to a reduction in m and E. The reduction in m increases k_m, whereas the reduction in E leads to a reduction in k_m. In total, k_m increases.

T_W **versus** T_K From (18a) and (18b), it follows that

$$1 + k_s^1 = k_m^1 - \hat{\gamma} M\hat{R}S_c^{21}\left(1 - \frac{k_m^1}{k_m^{21}}\right), \tag{24}$$

where $\hat{\gamma} = \frac{\gamma\psi_1'}{n^1\lambda_2} > 0$. The individual necessary conditions for s and m with the tax functions in (19) are

$$MRS_c^1 = k_m^1 - k_m^1(T_W^1 + T_K^1) = 1 + k_s^1 - (1 + k_s^1)T_W^1 - k_s^1 T_K^1.$$

The total distortion on m exceeds that on s, i.e

$$k_m^1(T_W^1 + T_K^1) > (1 + k_s^1)(T_W^1 + T_K^1) - T_K^1.$$

If $k_m^1 = 1 + k_s^1$, as in the first-best, then it would follow that $T_K^1 > 0$, and the sign of T_W^1 where ambiguous. Optimal allocations are however such that $k_m^1 > 1 + k_s^1$, then the sign of both T_W^1 and T_K^1 are ambiguous. However, since the capital income tax hits m harder than the wealth tax, one would expect $T_K^1 > 0$, though it cannot be shown in general.

If the tax function were changed such that the wealth tax base only consists of savings only, i.e. $T(s, k(E, s, m, a))$, then the marginal tax rate on capital income is unambiguously positive, and the marginal wealth tax rate is positive if T_K^1 is sufficiently large. Individual's first order conditions are

$$MRS_c^1 = k_m^1 - T_W^1 k_m^1 = 1 + k_s^1 - (1 + k_s^1) - T_W^1 - k_s^1 T_K^1.$$

Using the last equation and (24):

$$T_W^1 = T_K^1 - (1 - T_K^1)\hat{\gamma} M\hat{R}S_c^{21}\left(1 - \frac{k_m^1}{k_m^{21}}\right).$$