MEMORANDUM

No 06/2016

Optimal Taxation with Endogenous Return to Capital

Arnaldur Sölvi Kristjánsson

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 05/16	Øystein Kravdal *Expected and unexpected consequences of childbearing – a methodologically and politically important distinction that is overlooked*
No 04/16	Dag Fjeld Edvardsen, Finn R. Førsund and Sverre A. C. Kittelsen *Productivity Development of Norwegian Institutions of Higher Education 2004 – 2013*
No 03/16	Finn R. Førsund *Multi-equation modelling of Desirable and Undesirable Outputs Satisfying the Material Balance*
No 02/16	Ingrid Hjort *Potential Climate Risks in Financial Markets: Report from a workshop, January 20, 2016*
No 01/16	Ingrid Hjort *Potential Climate Risks in Financial Markets: A Literature Overview*
No 22/15	Geir B. Asheim and Frikk Nesje *Destructive intergenerational altruism*
No 21/15	Rolf Golombek, Alfonso A. Irarrazabal, Lin Ma *OPEC's market power: An Empirical Dominant Firm Model for the Oil Market*
No 20/15	Moritz A. Drupp, Mark C. Freeman, Ben Groom and Frikk Nesje *Discounting Disentangled*
No 19/15	Simen Markussen and Marte Strøm *The Effects of Motherhood*
No 18/15	Marit Linnea Gjelsvik, Ragnard Nymoen, and Victoria Sparrman *Have Inflation Targeting and EU Labour Immigration Changed the System of Wage Formation in Norway*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

Optimal Taxation with Endogenous Return on Capital*

Arnaldur Sölvi Kristjánsson†

Memo 06/2016

First version May 2016, this version August 2017

Abstract

This paper characterizes the optimal income and wealth tax schedules when rates of return are affected by effort and ability. Agents are heterogeneous along two dimensions: investment ability and labour market productivity. I show that when individuals can exert investment effort, the Atkinson-Stiglitz theorem that capital income should not be taxed does not hold. 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. When wealth is not observed, the optimal marginal tax rate on capital income is positive. The optimal marginal tax rates on labour and capital income should not be equal, but they are positively related to each other. As the 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. The results extend to a model where individuals can hire investment advisors to increase the rate of return, in which case the marginal tax rate on wealth may be positive. In addition, the results partly remain in a model with a domestic credit market.

JEL-codes: G11, H21, H24

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

*This paper is part of the research of Oslo Fiscal Studies supported by The Research Council of Norway. 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 Oslo. E-mail: kristjansson07@gmail.com

1 Introduction

In *Capital in the Twenty-First Century,* Thomas Piketty argues that the rate of return on capital increases with initial endowment because the rich can spend more on financial advisors. According to Piketty, this "can potentially give rise to a global dynamic of accumulation and distribution of wealth characterized by explosive trajectories and uncontrolled inegalitarian spirals. [...] only a progressive tax on capital can effectively impede such a dynamic" (Piketty, 2014: 439). His proposals are partly based on the model by Piketty and Saez (2013), where inequality is two-dimensional. Individuals differ both in their labour market productivity and in their inherited wealth. The optimal tax system is therefore two-dimensional, with a progressive tax on labour income and a progressive tax on inheritance. A missing piece in this analysis is the effects of heterogeneous returns on optimal taxation, which is explored in this paper.

The classical treatment of savings in economics assumes that individuals can save and borrow whatever amount they wish at an exogenously given interest rate. In such an environment, capital income reflects the shift of consumption between periods. In reality, the rate of return on savings differs among individuals. When individuals spend effort managing their portfolio, and thereby act as investors, capital income reflects savings and the return on the investment effort and ability. Therefore, capital income provides the government with information on individual's underlying skill level. If the government values redistribution from the skilled to the less skilled, this is a rationale for taxing capital income.

A well known result in optimal taxation from a two period extension of the Atkinson and Stiglitz (1976) model is that intertemporal allocations should not be distorted, i.e. capital income should not be taxed.[1] This result holds when available tax instruments include nonlinear labour income tax and preferences are weakly separable between labour and consumption. Nonlinear labour income tax is sufficient to raise revenue and redistribute resources because the intertemporal allocation only depends on income and not on individual's underlying productivity. Distorting the intertemporal allocation cannot distinguish individuals with different productivity beyond what the labour income tax does.[2]

The effect of heterogeneous returns, investment effort and investment ability is increasingly gaining attention in the literature.[3] Stiglitz (1985) sets up a simple

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

[2]The literature has considered a number of extensions of the Atkinson-Stiglitz model where the result of zero capital taxation does not hold. Including heterogeneous preferences, different initial wealth, presence of income shifting, uncertain future wages and borrowing constraint. See Banks and Diamond (2010) and Tuomala (2016) for reviews and references of these various arguments.

[3]Stiglitz notes that "[o]ne of the most important reasons for taxing capital income is that

model where individuals differ in the rate of return they obtain. Gerritsen et al. (2015) set up a model where individuals are endowed with heterogeneous abilities to earn labour income and to earn capital income. They show that capital income should be taxed if returns and labour market ability are positively correlated. The optimal capital income tax formula trades-off taxing capital income rents and distorting savings. The difference in my approach is that the return on capital is endogenous. Gahvari and Micheletto (2016) set up a related model, where the return is an increasing function of labour market productivity whereas it depends on effort and investment ability in my model. In the concluding section I shall further compare my work with theirs. Finally, Guvenen et al. (2017) set up a model with heterogeneous returns due to heterogeneous entrepreneurial productivity (but not on effort). They show that a shift from capital income to wealth taxation to be welfare improving, since this will lead to a more efficient allocation of capital.

The assumption that everybody faces the same interest rate is clearly unrealistic. Broadly speaking, there are two main explanations for why rates of returns are heterogeneous. The first explanation is that investors have asymmetric information on which investment options are likely to be good. The asymmetry follows because there is a fixed cost of acquiring information on which investment options are good, because it takes time to do market research and resources to collect information on investment options. Just as individuals differ in their labour market productivity, individuals could also differ in their ability to pursue market research. And secondly, differential rates of return may be due to uncertainty, which is an inherent feature of financial markets.

There is a growing literature studying determinants of capital returns. Two recent papers using Scandinavian administrative data find considerable heterogeneity in returns. Fagereng et al. (2016) study 20 years of Norwegian panel data and find that return differentials can largely be accounted for by persistent individual specific factors. Whereas the heterogeneity is only partly driven by differences in risk taking. Bach et al. (2015) analyse the relationship between returns and wealth using Swedish panel data. They find a strong correlation between returns and wealth, which is primarily driven by differences in risk taking. Differences in risk adjusted returns between wealth groups are significant but small.

There is also evidence from the US showing a positive correlation between returns and wealth for professional investors. Piketty (2014) shows that returns

we cannot clearly distinguish capital income from wage income, particularly the labor that goes into managing capital. When an investor gets an above average return, should the difference be viewed as a return to his skill as an investment manager and, therefore, really be viewed as a return to labor?" (2015: 246). In a similar vein, Diamond and Saez note that "[a] straightforward argument for taxing capital is that it is often difficult to distinguish between capital and labour incomes. For example, people spending time to manage their investment portfolios are converting labour time into anticipated capital income." (2011: 181).

on the endowments of US universities increases rapidly with the size of endowment. Interestingly, the volatility of returns is not related to endowment so that returns are systematically related to endowment. This indicates that higher returns are not primarily due to more risk taking, but rather due to a more sophisticated investment strategy. Saez and Zucman (2016) show that the same pattern emerges for the universe of U.S. foundations.[4] Interestingly, wealthy foundations have portfolios that are similar to rich families. Another piece of evidence from the Forbes global wealth rankings suggests that wealthier individuals tend to get higher returns (Piketty, 2014).

There is a growing literature on household finance and financial literacy that reveals the importance of financial literacy on financial outcomes. The literature has established a positive effect from financial literacy on important outcomes such as planning for retirement, stock market participation, diversification and reduced high cost borrowing.[5] The literature has pointed out four types of explanations for under-diversification, an important investment mistake. One of them reflects frictions in the portfolio choice (the other being financial literacy, behavioural biases and preferences for certain assets). Due to transaction and search costs individuals rationally choose to under-diversify as individuals trade-off costs and benefits (see review in Guiso and Jappelli, 2008). A striking result from Deuflhard et al. (2015) is that returns on savings accounts are considerably heterogeneous in Netherlands, which are very widely held and virtually riskless assets. Their results suggest that lack of information explains why households do not choose options with the highest returns. They argue that financial literacy affects the ability to identify accounts offering the highest return. This indicates the importance of getting information on investment options.

I set up a two period model of saving where individuals can exert investment effort, which increases the rate of return, and they work in the labour market. Individuals differ in investment abilities and labour market productivity. The government's aim is to redistribute resources from the skilled to the less skilled. I discuss whether the existence of different investment abilities and the possibility to make investment effort is a rationale to tax capital income. Preferences are separable between consumption and leisure and thereby satisfy the condition for the Atkinson-Stiglitz theorem. I show that a zero tax on capital income is not optimal when one introduces the possibility to exert investment effort when individuals

[4]The correlation is mainly due to the fact that unrealized capital gains rise with wealth, while realized returns are flat within asset classes.

[5]For literature reviews see Hastings et al. (2013), Lusardi and Mitchell (2014) and Stolper and Walter (2017). Whether the correlation can interpreted causally is challenging but both instrumental variables and experimental approaches suggest that literacy causally affects financial outcomes (and not the other way around).

differ either in labour market productivity or in investment abilities.

In my baseline model, there are two informational assumptions. In the first case, the government observes wealth and capital income. Here, the government wants to distort the investment decision downwards while keeping the savings decision undistorted. This is achieved by taxing capital income and subsidizing wealth (i.e. a negative marginal tax rate). The choice of investment effort and labour supply depends on skill since, conditional on income, more skilled individuals will have more leisure. Therefore, capital income provides the government with information on individual's underlying skill level and should be used for taxation. The intertemporal allocation that individuals choose does on the other hand not depend on skills. Therefore distorting the intertemporal allocation cannot distinguish individuals with different skill level.

In the second case, wealth is not observed whereas capital income is observed. Here, the intertemporal allocation is distorted by the capital income tax. The intuition for taxing capital income is similar to the first case. Conditional on income, more skilled individuals have a higher rate of return and save less and are therefore, at the margin, more willing to save. This means that capital income depends on individual's underlying skill level and should therefore be taxed.

The results from both cases show that the marginal tax rate on labour and capital income should not be equal, but they should be positively related to each other. In a model with heterogeneous labour market productivity (investment ability), the marginal tax rate on labour income exceeds (falls short of) the rate on capital income.

I make two extensions to the baseline model. In the first extension, I examine a model where individuals can hire financial advisors to increase their rate of return. In this extension, the government both wants to distort capital income as well as the intertemporal allocation, i.e. tax capital income and wealth. In the second extension I add a domestic credit market allowing for interpersonal lending. Such transactions would be Pareto improving. In this model, the sign of the optimal capital income tax is ambiguous. There are firstly the effects from the baseline model calling for a positive tax. Secondly, the capital income tax rate will affect the interest rate on the domestic credit market and if a reduction in the interest rate is welfare improving this calls for a positive capital income tax. Whether a decrease in the interest rate is beneficial or not from the government's point of view, depends on who is the borrower and who is the lender.

This paper is organized as follows. Section 2 sets out the model and section 3 solves the full information benchmark (i.e. the first best). In sections 4 and 5 I solve the second best problem with different informational assumptions. I extend the model in various directions in section 6 and conclude in the final section.

2 The model

Following Mirrlees (1971) optimal income tax models generally 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, the 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 is exogenously given. Individuals are born with them and have no opportunities to affect them. Apart from this, individuals are identical in all other respect. There is for example no preference heterogeneity.

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.

Each agent supplies L units of labour in the first period. The labour market is perfectly competitive and individuals of different productivity are perfect substitutes so workers receive a fixed wage of w based on their exogenously given ability. Labour income is denoted as $Y = wL$.

Time spent investing increases the return on savings. The time spent can be thought of as market research, where individuals learn about the profitability of investment projects. Investments are made in a set of existing investment projects. Investment effort constitutes of spending time in finding good investment project (i.e. that have a high return). It should be emphasized that individuals are investors only by managing their own portfolio and not the portfolio of other people.

The economy is small and open and individuals pursue investments in an international investment market. Since the economy is small and open, the behaviour of individuals will have no general equilibrium effects.

Capital income is denoted by $k(E, s, a)$ which is increasing in E, s and a, where s denotes savings and k_s is the return on savings. In addition, it has 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 $k_{Ea} > 0$ means that more able investors are more efficient, which is my definition of being an abler investor. The cross-derivative k_{sE} is positive, meaning that the rate of return is increasing in investment effort.

Both savings and investment effort have weakly decreasing returns. When $k_{ss} = k_{EE} = 0$, the capital income function could take the form $k(E, s, a) = Esa$. There is a given amount of heterogeneous investment options, where each investment option is of finite size. Therefore, for a given E and a, an increase in s will either decrease or keep k_s unchanged

The budget constraint of individuals in period 1 and 2 are

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

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 is increasing and concave in c_1, c_2 and leisure. The separability between leisure and consumption is in order to avoid the effects of complementarity on optimal taxation, an issue that has received great attention in the literature (see e.g. Christiansen, 1984). The total time available is normalized to 1. Individuals face a time constraint 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, it maximizes the sum of utilities. Since individuals have concave utility functions, the government has a redistributive motive. It is assumed that the government does not know individual skill level (w and a) nor individual labour supply and investment effort (L and E). The government observes both labour income and capital income at the individual level and knows the distribution of w and a and 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. Capital income usually comes in form of a transaction (e.g. dividends and interest) and is therefore easy to observe for tax authorities. Estimating the market value of assets is on the other hand a much more daunting exercise.[6] I believe that both informational assumptions are 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 income for every type in both periods. Then, the government solves the problem subject to the incentive constraints to prevent a certain type 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 and a total of four types of individuals. Labour market productivity can be either high or low, denoted by w^h and w^l, respectively, with $w^h > w^l$. By the same token, investment effort can be either high or low, denoted by a^h and a^l, respectively, with $a^h > a^l$.[7]

3 First best

In the first best, the government knows the investment ability and labour market productivity of all individuals and thereby it is not concerned with an incentive constraint. The only constraint that the government faces is a budget constraint. The government's problem is to maximize the sum of utilities

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

where $I = 4$, n^i denotes the number of individuals of type i, $B_1^i = Y^i - t^i$ denotes disposable income in the first period for individual i (or post tax labour income), which can be spent on consumption and savings. $B_2^i = k(E^i, s^i, a^i) - T^i$ denotes disposable income in the second period for individual i (or post tax capital income). In (2), 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 period 1 and 2, denoted

[6] Slemrod and Gillitzer (2014) argue that basing tax liability on market transactions has several advantages. They say that "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).

[7] In the optimal tax literature, the standard assumption to make is that individuals only differ in terms of labour market productivity. Many examples exist of models with multiple heterogeneity, see e.g. Cremer et al. (2004).

by g_1 and g_2, respectively. This can be interpreted as required revenue for essential public goods. For simplicity, the government is not allowed to borrow or save between periods. Importantly this will not affect the main qualitative results that are derived.[8] The government's budget constraint, and also the resource constraint, in period 1 and 2 are, respectively

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

λ_1 and λ_2 denote the multipliers associated to the budget constraint in period 1 and 2, respectively. Necessary conditions for $i = 1, 2, 3, 4$ are

$$\begin{aligned}
&(s^i) && -u_i' + \psi_i' + \lambda_2 k_s^i = 0,\\
&(E^i) && -v_i' + \lambda_2 k_E^i = 0,\\
&(Y^i) && -v_i'/w^i + \lambda_1 = 0,\\
&(B_1^i) && u_i' - \lambda_1 = 0,\\
&(B_2^i) && \psi_i' - \lambda_2 = 0.
\end{aligned}$$

Partial derivatives of the functions u, ψ and v are denoted with a prime and the subscript indicates the corresponding type, e.g. $u_i' = \frac{\partial u(c_1^i)}{\partial c_1^i}$. As is standard in first best problems like this, the government aims at equalizing marginal utility of consumption in both periods and minimizes the utility loss of effort. This means that everybody will have the same consumption in both periods. Since more able individuals are more efficient in their time use, individuals with higher w will supply more L and by the same token individuals with higher a will supply more E.

By eliminating Lagrange multipliers, the optimal intertemporal and intratemporal allocations can be presented as

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

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

[8]This means though that the timing of taxation matters. In other words, the Ricardian equivalence does not hold. If I would allow for government borrowing/saving, the Ricardian equivalence would on the other hand hold. Government borrowing/saving could be added without affecting the main qualitative results but would nonetheless affect the optimal intertemporal allocation. The effects of government borrowing largely depend on the interest rate that the government faces. If the government faces a large interest rate, the government would perform investments by imposing large taxes in period 1 and low taxes in period 2. If the government faces a low interest rate, the government would borrow and let individuals perform investments with the borrowed money.

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

where MRS_Y^i denotes the marginal rate of substitution between labour income and present consumption. It shows how much an individual needs to be compensated in terms of present consumption when supplying one more unit of labour income in order to be indifferent. 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 the marginal rates of transformation, which is 1, and (4c) shows that the intertemporal marginal rate of substitution should be equal to $1 + k_s^i$.

Combining (4a)-(4c), it follows that

$$k_E^i = w^i(1 + k_s^i),$$

which indicates that the marginal return on time use, in present value terms, is equated between L and E.

4 Government observes Wealth

I now derive the optimal allocation subject to the government being information and resource constrained. The government has information on individual's labour income, savings and capital income. This means that the government knows individual capital income as well as their wealth. Capital income that is assigned by the government is denoted by K while $k(E, s, a)$ is the amount of capital income received by the individual, where these two have to be equated, $K = k(E, s, a)$.

To avoid difficulties with multidimensional screening I will not consider the case of all four possible types of individuals and only consider only a two type model (a four type model is considered in section 5.4). First, I consider a two type model with fixed a where individuals differ in w, with $w^2 > w^1$. Second, I consider a two type model with fixed w where individuals differ in a, with $a^2 > a^1$. In both models the objective of the government is (2), the same as in the first best, subject 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 type 2 individual choosing the bundle for type 1 and $\hat{E}^{21}$ is the investment effort chosen by a type 2 person mimicking a type 1 person, it is the value of E such that $K^1 = k(\hat{E}^{21}, s^1, a^2)$. (5) ensures that a type 2 individual does not choose the bundle intended for a type 1 individual.

An important feature of the model is that mimickers have more leisure than the type being mimicked. This is because mimickers with a higher w will supply less labour while mimickers with a higher a will exert less investment effort, hence $L^1 + E^1 > \hat{L}^{21} + \hat{E}^{21}$.

4.1 Two Type Model: fixed Investment Ability

Here, I consider a two type model ($I = 2$) with fixed a where individuals differ in w, with $w^2 > w^1$. In appendix A, the Lagrangian is presented, the necessary conditions derived and manipulated. The optimal allocation is

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

$$MRS_K^2 = 1, \tag{6b}$$

$$MRS_c^i = 1 + k_s^i = \frac{\lambda_1}{\lambda_2}, \quad i = 1, 2, \tag{6c}$$

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

$$MRS_Y^2 = 1, \tag{6e}$$

where γ is the Lagrange multiplier associated with the incentive constraint, $M\hat{R}S_K^{21} = \frac{\hat{v}_{21}'}{\psi_1'} \frac{1}{\hat{k}_E^{21}}$ 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). These conditions indicate that type 2 individual should be left undistorted. If implemented with a tax system, type 2 faces a marginal tax rate of zero. This is the standard no distortion at the top result as in the Mirrlees model, which has received great attention in the literature, and will hold in all applications that are considered, except for the model in section 6.3. Therefore, it hardly needs further explanation.

Conditions (6a) and (6d) show that capital income and labour income should be distorted downwards, i.e. investment effort and labour supply are distorted downwards. According to (6c) the intertemporal allocation of type 1 should though be left undistorted. This means that the rate of return (k_s^i) will be constant across agents. It should be emphasized that in the model considered here both types have the same investment ability.

Implementing the above allocation with a tax system will put a positive marginal tax rate on labour and capital income. Taxing capital income will distort both investment effort and the intertemporal allocation. In order to keep the intertemporal allocation undistorted, savings should be distorted upwards. There will be a positive marginal tax rate on capital income but a negative marginal tax rate on wealth. As is shown below, the marginal tax rate on labour income exceeds the rate on capital income. However, the marginal tax rates are positively related to each other (see appendix A for the formal derivations).

Individuals with higher w have, conditional on labour income, more leisure because their labour supply is lower. Therefore, individuals that are more productive are more willing to supply more labour, conditional on income. Therefore, labour income provides the government with information on productivity. Since the government wants to redistribute from the high productive to the less productive they should use labour income for taxation. Capital income and the choice of investment effort does not depend on labour market productivity. However, more productive individuals have more leisure, conditional on income, and they are therefore at the margin more willing to exert investment effort. This provides the government with information about individual's underlying productivity and should be used for taxation. The intertemporal allocation is however independent of productivity; it only depends on income. Therefore distorting the intertemporal allocation cannot distinct individuals with different productivity beyond that what the labour and capital income does.

Capital income and labour income provide 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. Therefore, labour income and capital income should be used for taxation purposes. In addition to this, individuals that are more productive are, conditional on income, more willing to supply more labour because they are more productive. This is a further argument for distorting labour income. Therefore, labour income should have a higher marginal tax rate than capital income, but the marginal tax rates should be positively related to each other.

In the first best, the government completely equalizes consumption but in the second best the government is restrained from doing so because of the incentive constraint, which is binding at the optimum. The reason that the government wishes to tax capital income is that this serves to relax the binding incentive constraint and thereby the government can achieve a Pareto improvement. This is possible because a type 2 mimicker has a lower marginal rate of substitution than type 1 individual, i.e. $MRS_K^1 > M\hat{R}S_K^{21}$.[9] A type 2 mimicker faces the same bundle as type 1 but needs less compensation in terms of future consumption to supply one more unit of capital income because the mimicker has more leisure and is

[9]Note that at the optimum no one will mimic, there are only potential mimickers. But the behaviour of the mimicker determines how far the government can go in redistribution.

therefore more willing to give up leisure (compared to the less productive worker). This means that the less productive worker has a steeper indifference curve in the K, c_2 space (see figure 1). This can be explained by performing a perturbation. Suppose one starts from an undistorted allocation which satisfies the incentive constraint, where $MRS_K^1 = 1$. Consider a small variation of $dK^1 < 0$ with a variation of $dB_2^1 \cdot MRS_K^1 = dB_2^1 = dK^1$. This small variation is simply a small change along types 1 indifferent curve and has therefore no effect on the utility of type 1. But the mimicker has a steeper indifference curve and is therefore not at an undistorted allocation ($M\hat{R}S_K^{21} < 1$ when $MRS_K^1 = 1$). This small variation will therefore decrease the utility of type 2 mimicking type 1 while type 1 is indifferent. In other words, the downward distortion will make mimicking less attractive and therefore relax the incentive constraint.

The optimal allocation for type 2 and 1 are shown graphically in figure 1, which closely resembles the static Mirrlees model. The more productive worker is located at point B, where the investment effort is undistorted. Their indifference curves cross at point A, which is the allocation of the less productive worker who is distorted downwards. The figure shows that the more able investor will have a higher capital income and a higher consumption in period 2, i.e. $K^2 > K^1$ and $c_2^2 > c_2^1$.

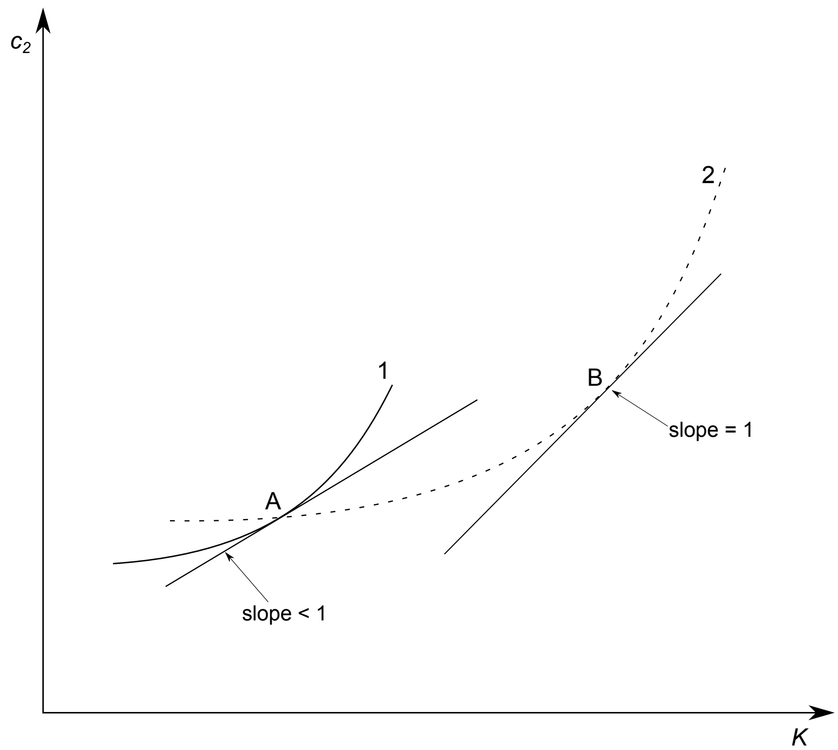

Figure 1: Indifference curves in K and c_2 space under optimal allocation.

It has been established that capital income should be distorted downwards while the intertemporal allocation should not be distorted. To show how that would be implemented by a tax system, let me consider the individual's budget constraint with tax functions in both periods

$$c_1 = Y - s - t(Y),$$
$$c_2 = s + k(E, s, a) - T(s, k(E, s, a)).$$

From conditions (6a) and (6c) and the individual's necessary conditions, it follows that t', $T_K > 0$ and $T_s < 0$. [10] This means that the marginal tax rate on capital income is positive while the marginal tax rate on savings will be negative. A positive tax on capital income will distort investment effort as well as the intertemporal allocation. In order for the intertemporal allocation to be undistorted, savings are subsidized at the margin. That is, the marginal subsidy of savings ensures that the intertemporal allocation will be undistorted.

4.2 Two Type Model: fixed Labour Market Productivity

Here, I consider a two type model with fixed w where individuals differ in a, with $a^2 > a^1$. The optimal allocation that results in this model is analogous to equations (6a)-(6e) and no need to repeat them. The distortions that result are qualitatively the same as above and the same inequalities as in (6a)-(6e) apply. Labour income and capital income of the less able investor is distorted downwards while the intertemporal allocation is left undistorted.

The downward distortion on capital income should not come as a surprise. I have treated capital income very similar to labour income in the Mirrlees model and therefore it should be natural to distort capital income downwards. The reason for this is the same as in the model above, it serves to relax the incentive constraint (the same discussion as above applies).

Condition (6d) shows that labour supply of the less able investor should be distorted downwards, just as in the model with fixed a. Remarkably, I get this result also here where both types have the same labour market productivity.

Opposite to the results from section 4.1, the marginal tax rate on capital income will be higher than for labour income, i.e. $t' < T_K$. The intuition is exactly opposite to the arguments in section 4.1. Firstly, labour and capital income are distorted because mimickers are at the margin more willing to work because they have more leisure time. Secondly, mimickers are even more willing to increase investment effort than labour supply because mimickers are more skilled investors (while they are equally productive workers). The results from this section are summarized in proposition 1.

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

Proposition 1 *If the government observes wealth, capital and labour income, the optimal tax system has a positive marginal tax rate on capital and labour income but a negative marginal tax rate on wealth. This will distort investment effort while the intertemporal allocation undistorted. The marginal tax rate on labour income will exceed (fall short of) the rate on capital income if people differ in terms of labour market productivity (investment ability).*

5 Government does not observe Wealth

5.1 Comparative Statics of Individuals

Here, I consider the case where the government only observes labour income and capital income and knows the distribution of w and a as well as preferences. The government does not observe s^i. Therefore, if an individual reports high capital income, the government does not know to what extent that is due to savings, investment ability or investment effort.

The government offers individuals bundles in terms of pre- and post-tax income in both periods, which is the bundle (Y, B_1, K, B_2). Individuals choose among the bundles offered by the government.

Given the bundle that individuals choose, they have no degree of freedom w.r.t. L. In order to produce a given level of Y, individuals set labour supply as $L = Y/w$. Regarding the bundle in the second period, individuals have one degree of freedom. Individuals choose both E and s but are constrained by the fact that capital income needs to equate the level of K set by the government, or

$$k(E, a, s) = K, \tag{7}$$

where K is the quantity chosen by the government and $k(E, a, s)$ is capital income that individuals receive. Instead of performing a constrained maximization, individuals choose E freely and let s adjust according to the constraint. This implicitly defines $s = s(E, a, K)$ by (7).

Now an individual who is offered the bundle (Y, B_1, K, B_2) faces the following problem

$$\max_{\{E\}} U = u(B_1 - s(E, a, K)) + \psi(B_2 + s(E, a, K)) + v(1 - Y/w - E). \tag{8}$$

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

$$U_E = -u' s_E + \psi' s_E - v' = 0, \tag{9}$$

where $s_E = -k_E/k_s < 0$. Condition (9) shows that a small increase in E leads to an increase in first period consumption (since savings are reduced, due to (7)), a decrease in second period consumption and a decrease in leisure.

From the necessary conditions in (9), it follows that the optimal choice of E is a function of all the exogenous variables, i.e. $E = E(Y, B_1, K, B_2, w, a)$. By implicitly differentiating (9), I can derive the derivatives of E w.r.t. w and a in order to analyse the behaviour of the mimicker,

$$\frac{dE}{dw} = \frac{-U_{Ew}}{U_{EE}} = \frac{v''}{-(u'-\psi')s_{EE} + (\psi''+u'')s_E^2 + v''}\frac{Y}{w^2} > 0, \tag{10}$$

$$\frac{dE}{da} = \frac{-U_{Ea}}{U_{EE}} = \frac{(u'-\psi')s_{Ea} - (\psi''+u'')s_E s_a}{-(u'-\psi')s_{EE} + (\psi''+u'')s_E^2 + v''} < 0, \tag{11}$$

where $s_a = -k_a/k_s < 0$ and $s_{EE} = (k_{ss}s_E k_E + 2k_{sE}k_E - k_{EE}k_s)/k_s^2 > 0$. Note that in (10) and (11), 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 (10) 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 that has a higher w have more leisure than the less able worker. It follows from (11) that mimickers who have higher a, will also have more leisure than the type being mimicked.

Having determined how the optimal choice of E changes with w and a, I next explore how s changes with w and a, again, conditional on Y, B_1, K and B_2. This is done by differentiating $s(E(Y, B_1, K, B_2, w, a), a, K)$

$$\frac{ds}{dw} = s_E\frac{dE}{dw} < 0, \tag{12}$$

$$\frac{ds}{da} = s_E\frac{dE}{da} + s_a < 0, \tag{13}$$

see appendix B for the sign of the inequality in (13), which can be shown by some manipulation.

When labour market productivity increases, individuals can provide a given level of Y at a lower L. Then total utility of leisure (which is $1 - L - E$) increases. Since the utility function is concave, individuals respond by increasing E and decreasing s. This means that a mimicker will choose to save less and exert more investment effort than the type being mimicked, i.e. $\hat{s}^{ji} < s^i$ and $\hat{E}^{ji} > E^i$. $\hat{s}^{ji}$ and $\hat{E}^{ji}$ denote the saving and investment effort, respectively, of a type j mimicking type i individual. Due to the a higher labour market productivity, a mimicker will mechanically supply less labour, $\hat{L}^{ji} < L^i$. In total, the mimicker will have more leisure.

When the investment ability increases, an individual needs to exert less investment effort and save less to provide a given K. Therefore, he will find it beneficial to decrease investment effort and savings. This means that the high ability mimicker will choose to save less and exert less investment effort, i.e. $\hat{s}^{ji} < s^i$ and $\hat{E}^{ji} < E^i$.

In order to formulate the indirect utility function, I plug the optimal $E(\cdot)$ into the utility function. The indirect utility function is denoted by $V(Y, B_1, K, B_2)$. The derivatives of the indirect utility function follow 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 $s_K = 1/k_s$. Note that this holds for individuals choosing the bundle intended for them as well as for mimickers.

5.2 Two Type Model: fixed Investment Ability

I now derive the optimal allocation subject to the government being information and resource constrained. To avoid difficulties with multidimensional screening I will, to begin with, not consider the case of all four possible types of individuals. I consider a two type model, where individuals only differ in terms of their labour market productivity while investment ability is fixed. Thereby I am staying as close as possible to the Atkinson-Stiglitz environment. Type two has a higher labour market productivity, $w^2 > w^1$. The allocation is chosen such that a type 2 individual does not choose the bundle intended for type 1. The government offers bundles in terms of Y, B_1, K, B_2 for both types. As before, the government's problem is to maximize the sum of utilities

$$\begin{aligned} \max_{\{Y^i, B_1^i, K^i, B_2^i\}} \quad & \sum_i^2 n^i V^i \\ \text{subject to} \quad & \sum_i^2 n^i (Y^i - B_1^i) \geq g_1 \quad (\lambda_1), \\ & \sum_i^2 n^i (K^i - B_2^i) \geq g_2 \quad (\lambda_2), \\ & V^2 \geq \hat{V}^{21} \quad (\gamma). \end{aligned} \tag{14}$$

I denote $\hat{V}^{21} = V(Y^1, B_1^1, K^1, B_2^1, w^2, a^2)$ as the indirect utility of a type 2 person mimicking 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 him. At the optimum, the above constraints hold with equality. The corresponding Lagrange multipliers are shown in parenthesis in (14).

In appendix B I set up the Lagrangian, derive the necessary conditions and manipulate them. The optimal allocation is

$$MRS_c^1 =(1+k_s^1) - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\Big[(MRS_c^1 - M\hat{R}S_c^{21}) + (1-k_s^1/\hat{k}_s^{21})(M\hat{R}S_c^{21}-1)\Big] < 1+k_s^1, \tag{15a}$$

$$MRS_Y^1 =1 - \frac{\gamma\hat{u}'_{21}}{n^1\lambda_1}\Big[MRS_Y^1 - M\hat{R}S_Y^{21}\Big] \lesseqgtr 1. \tag{15b}$$

Condition (15a) shows that savings of type 1 should be distorted downwards. If implemented with taxes, then at the optimum there should be a positive marginal tax on capital income. This is quite a remarkable result. The only thing I have added to the standard two period model that lead 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. It should be noted that conditions (15a) is contingent on individuals exerting investment effort. If nobody exerts investment effort (and there is a corner solution), then savings should not be distorted.

The government wants to redistribute from the more productive to the less productive. More productive individuals will, conditional on income, choose a different intertemporal allocation (they will front load consumption) and they are therefore at the margin more willing to save. This means that the intertemporal allocation provides the government with information on labour market productivity 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 as explained below. This reflects the first term in the bracket in (15a). Second, a mimicker has a higher rate of return than the type being mimicked, i.e. $k_s^i < \hat{k}_s^{ji}$. This reflects of the second term in the bracket in (15a). Both effects imply that mimickers are more willing to increase savings. Therefore distorting savings downwards, makes mimicking less attractive.

As discussed above, a type 1 individual saves more than the mimicker (i.e. $s^1 > \hat{s}^{21}$). Therefore a less productive worker has a higher MRS_c than the mimicker, i.e. $MRS_c^1 > M\hat{R}S_c^{21}$. Compared to the mimicker; a less productive worker needs to get a larger compensation in terms of future consumption in order to forego one unit of present consumption. This means that less productive types have a steeper indifference curve in the (c_1, c_2) space than a more productive type (see figure 2). In addition, a mimicker has a higher rate of return. This also implies that the mimicker needs less compensation in terms of future consumption in order to forego one unit of present consumption.

The optimal allocation for type 1 and 2 are shown graphically in figure 2, which resembles closely the static Mirrleesian model.[11] The productive worker is located at point B, where the intertemporal allocation is left undistorted. Their indifference curves cross at point A, which is the allocation of the less productive worker who is distorted downwards.

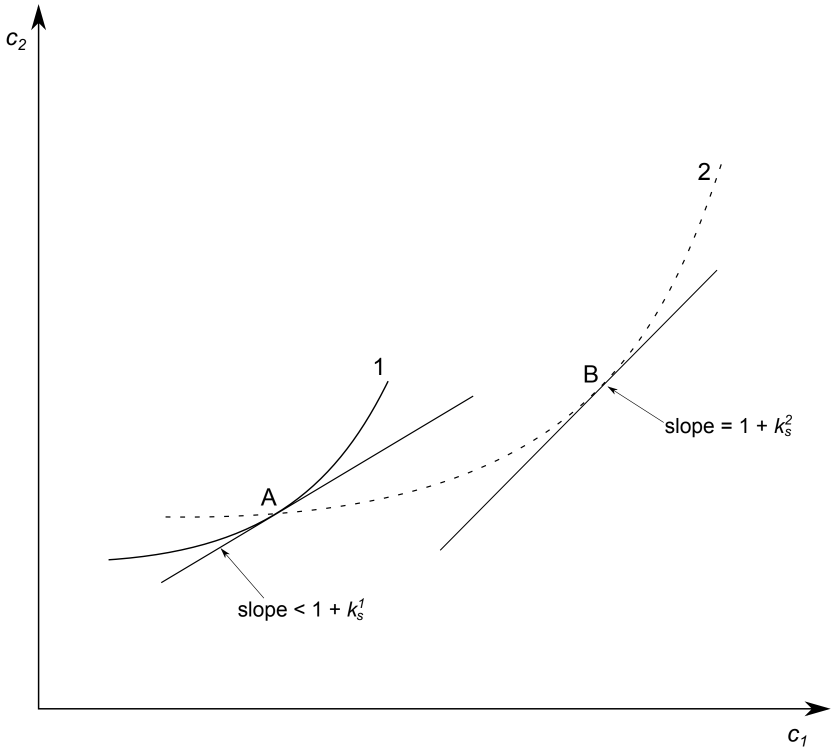

Figure 2: Indifference curves in c_1 and c_2 space under optimal allocation.

Condition (15b) shows that the direction of the distortion, if any, on the labour-leisure decision of type 1 individual is ambiguous. The reason 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 less productive worker. 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 produce one more unit of Y. This calls for an upward distortion of 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 (15b) is ambiguous.

I want to know whether the distortion on labour income exceeds or falls short of the distortion on capital income. Since individuals receive labour income in

[11] The indifference curves show the combination of B_1, B_2 that holds $u(B_1 - s(E, a, K)) + \psi(B_2 + s(E, a, K)) + v(1 - Y/w - E)$ constant.

the first period and capital income in the second period, it is not possible to assess whether the distortion on labour income or capital income is greater as the intertemporal allocation is being distorted. In order to proceed, I change the setup such that labour income and capital income are both paid in the second period. Individuals have an homogeneous endowment e in period 1 and face the following budget constraint

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

In appendix B, I show that the marginal tax rate on labour income exceeds the marginal tax rate on capital income. The reason is analogous to the insights from section 4.1. Mimickers are more willing to increase investment effort and labour supply because they have more leisure time and therefore their marginal dis-utility of L and E are lower. In addition to this are mimickers even more willing to increase L than E because mimickers are more productive workers. Finally, mimickers are at the margin less productive investors (as $k_E^1 > \hat{k}_E^{21}$), and therefore they are less willing to increase E than L (see appendix B). In this model where labour income is received in the second period, labour supply is distorted downwards. That is, condition (15b) has $MRS_Y^1 < 1$.

The results suggest that marginal tax rates for labour and capital income should not be equal but they also show that the marginal rates should be positively related to each other, just as in section 4.1 (for this statement to hold I am keeping $\frac{w^1}{w^2}$ and $\frac{k_E^1}{w^1(1+k_s^1)}$ constant).

5.3 Two Type Model: fixed Labour Market Productivity

I analyse the same problem as in (14), except that now individuals differ in their investment ability (with $a^2 > a^1$) and have the same labour market productivity. The optimal allocation will have the same form as in (15a-15b). The rationale and intuition for the distortions are identical, they serve to relax the incentive constraint.

An important feature of the distortion on capital income from condition (15a) is that it does not depend on individuals exerting any investment effort. This means that even though no individual would exert any investment effort in this model, the government would tax capital income. What matters for the distortion is that mimickers save less then the less able investor. Whether type 1 has zero or positive E is not relevant for the result that capital income should be taxed.[12]

As in the model with heterogeneous w, there is a downward distortion on capital income. The rationale is identical to the model above. First, mimickers have a lower intertemporal marginal rate of substitution (since they save less). Second, a mimicker has a higher rate of return than the type being mimicked, i.e. $k_s^1 < \hat{k}_s^{21}$ since he is a more effective investor (see appendix B). Mimickers are due to these two differences more willing to increase savings and therefore a downward distortion will make mimicking less attractive.

Turning to the labour-leisure allocation for type 1 individuals in (15b), it is not possible to determine the sign of optimal distortion. The reason is that there are two counteracting forces (the same as in section 5.2.).

Finally, I turn to the difference in distortions between labour and capital income. As in section 5.2, it is uncertain whether the marginal tax rate will be higher for labour income than capital income. Following the same procedure as above and letting individuals receive labour and capital income in the second period, I show in appendix B that the marginal tax rate on capital income should exceed the marginal tax rate on labour income. This is exactly the opposite from the results in section 5.2. Firstly, mimickers are more willing to increase E or L because they have more leisure time. Secondly, mimickers are even more willing to increase E compared to L because mimickers are more able investors (while they are equally productive workers). The results are summarized in proposition 2.

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. When individuals receive labour and capital income in the same period, the marginal tax rate on labour income will exceed (fall short of) the rate on capital income if people differ in terms of labour market productivity (investment ability).*

5.4 Four Type Model

Here I analyse the general model with all four types of individuals. The distribution of types is shown in figure 3. 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

[12] If the less able type has $E^1 = 0$, the mimicker has to set $k(\cdot) = K^1$ such that:

$$k(0, a^1, s(0, a^1, K^1)) = k(\hat{E}^{21}, a^2, s(\hat{E}^{21}, a^2, K^1)).$$

At the interior, $\frac{dE}{da} < 0$, hence $\hat{E}^{21} = 0$ if $E^1 = 0$. It follows therefore from the above condition that $\hat{s}^{21} < s^1$, since $s_a < 0$.

in figure 3. 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 possibly 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 3 which shows all the possibly binding incentive constraints. It is not possible to rule any of them out a priori. Since I do not have a single-crossing property, it is not possible to rule out global incentive constraints. The single-crossing property is satisfied for the indifference curve in the c_1, c_2 space, since $\frac{\partial MRS_c}{\partial w}, \frac{\partial MRS_c}{\partial a} < 0$, but it is not satisfied for the indifference curve in the Y, c_1 space (see appendix C).

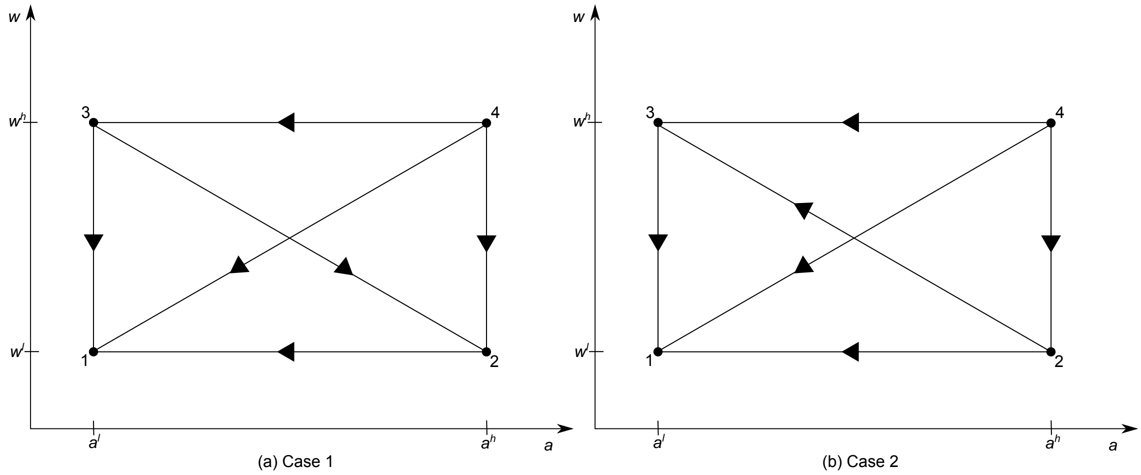

Figure 3: Potentially binding incentive constraints in case 1 and 2.

I can reveal the following general results: In case 1(2), type 2(3) should have a positive marginal tax rate on capital income while type 3(2) faces an ambiguous distortion. Type 1 should always have a positive marginal tax rate on capital income (see appendix C). The direction of the ambiguous distortion depends on the joint distribution of a and w.

When it is known which case applies, it is ambiguous how the mimicker will behave in comparison to the type being mimicked. Therefore, the sign of the distortion on either type 2 or type 3 are ambiguous. The reason is that it cannot be shown in general whether type 3(2) mimicking type 2(3) will save more, and who has a higher rate of return. This is because there are two opposing forces. Mimickers who have either higher w or a, will save more. Since type 3 has a high w and low a than type 2, it is not clear which effect will dominate.

In what proceeds I define $\Delta = w^h - w^l$. To make the result more clear, I will fix $a^h - a^l$ and vary Δ. If $\Delta = 0$ then clearly $V^2 > V^3$ and case 2 applies. Also, it is clear due the comparative statics from section 5.1 that $s^3 > \hat{s}^{23}$ and that $k_s^3 < \hat{k}_s^{23}$. Then, the optimal tax system distorts the intertemporal allocation of types 2 and 3 downwards. As Δ increases, there will be three critical values of Δ:

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

As long as the capital income tax function is separable such that $k(E, a, s) = g(E, a)f(s)$, then $\Delta = \widetilde{\Delta}$ also implies that $k_s^2 = \hat{k}_s^{32}$ and $\Delta = \bar{\Delta}$ implies $k_s^3 = \hat{k}_s^{23}$, which I assume to be the case.[13] This is important because then $\Delta = \widetilde{\Delta}$ implies $MRS_c^2 = \hat{MRS}_c^{32}$ and $k_s^2 = \hat{k}_s^{32}$ while $\Delta > \widetilde{\Delta}$ implies $MRS_c^2 > \hat{MRS}_c^{32}$ and $k_s^2 < \hat{k}_s^{32}$. Likewise, $\Delta = \bar{\Delta}$ implies $MRS_c^3 = \hat{MRS}_c^{23}$ and $k_s^3 = \hat{k}_s^{23}$ while $\Delta < \bar{\Delta}$ implies $MRS_c^3 > \hat{MRS}_c^{23}$ and $k_s^3 < \hat{k}_s^{23}$.

When $\Delta = \hat{\Delta}$, the effects of a high w for type 3 (implying $V^2 < V^3$) exactly outweighs the effects of a low a for type 3 (implying $V^2 > V^3$). When $\Delta = \widetilde{\Delta}$, a type 3 mimicking type 2 (i.e. choosing the bundle intended for type 2) will save the same as type 2. Then the effects of a high w (implying $s^2 > \hat{s}^{32}$) will exactly cancel out the effects of a low a (implying $s^2 < \hat{s}^{32}$). Likewise, when $\Delta = \bar{\Delta}$, the effects of a low w (implying $s^3 < \hat{s}^{23}$) will exactly cancel out the effects of a high a (implying $s^3 > \hat{s}^{23}$). Intuitively, $\hat{\Delta}$, $\widetilde{\Delta}$ and $\bar{\Delta}$ should not be too different, or equal, as the size of the cut-off values are determined by the same two effects that outweighing each other (the effects of a low/high w and a).

In general, it cannot be shown whether $\hat{\Delta}$, $\widetilde{\Delta}$ and $\bar{\Delta}$ are equal or not. If $\hat{\Delta} = \widetilde{\Delta} = \bar{\Delta}$, it is clear that both type 2 and type 3 will face a downward sdistortion for all values of Δ. When $\Delta < \hat{\Delta}$ case 2 applies and then $MRS_c^3 > \hat{MRS}_c^{23}$ and $k_s^3 < \hat{k}_s^{23}$. Then a downward distortion on savings is optimal as it will relax the incentive constraint, for reasons that are discussed in section 5.2. When $\Delta > \hat{\Delta}$ case 1 applies and then $MRS_c^2 > \hat{MRS}_c^{32}$ and $k_s^2 < \hat{k}_s^{32}$ and a downward distortion on savings is optimal. In either case, all the three types will face a downward distortion.

Finally, when $\min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}) < \max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})$ there will be an ambiguous distortion on either type 2 or type 3 when $\Delta \in \Big(\min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}), \max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})\Big)$, unless when $\widetilde{\Delta} < \hat{\Delta} < \bar{\Delta}$. The range of Δ under which case 1 or 2 applies and the range where type 2 or 3 possibly faces an ambiguous distortion depends on the ordering of $\hat{\Delta}$, $\widetilde{\Delta}$ and $\bar{\Delta}$. Determining the regions is not in itself important, all that matters is that the smaller $\max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}) - \min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})$ is, the smaller is the

[13] By definition $g(E^2, a^2)f(s^2) = g(\hat{E}^{32}, a^3)f(\hat{s}^{32})$. Now if $s^3 = \hat{s}^{32}$, then $g(E^2, a^2) = g(\hat{E}^{32}, a^3)$ and hence $k_s^2 = g(E^2, a^2)f'(s^2) = \hat{k}_s^{32} = g(\hat{E}^{32}, a^3)f'(s^2)$. The same applies for $k_s^2 = \hat{k}_s^{32}$.

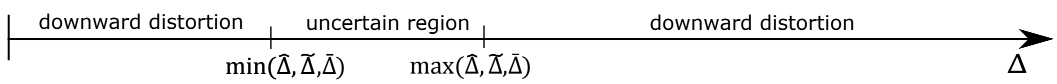

Figure 4: The direction of intertemporal distortion on type 2 and 3 when $\min(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta}) < \max(\hat{\Delta}, \widetilde{\Delta}, \bar{\Delta})$.

uncertain region depicted in figure 4. I summarize the main results of this section 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 4. If Δ is within the uncertain region, the marginal tax rate on type 2 or 3 may be ambiguous.*

6 Extensions

6.1 Buying Investment Information

In this section, I expand the model such that individuals can spend money to gain information which will give them a higher rate of return. Piketty (2014) argued this to be the main reason for why returns are heterogeneous. This can be thought of as hiring a financial advisor or employing a wealth management consultant. This will enable individuals to make better investments and get a higher rate of return. I do this by adopting the capital income function $k(m, \cdot)$, where m denotes the expenditure on financial advisory. There are three cases of the capital income function depending on which arguments are effective

1. $k(m, s)$: no investment effort
2. $k(m, s, a)$: no investment effort and investment ability matters
3. $k(E, m, s, a)$: investment effort possible and investment ability matters

The first case constitutes of an environment where individuals completely follow the financial advisor and individual investment ability does not matter. It is as if the financial advisor will completely manage your portfolio and the rate of return will be higher, the more is spent on financial advisors. In the second case investment ability matters. This can be thought of as an environment where investment effort is fixed for all individuals. The third case adds the possibility to exert investment effort as well as hiring a financial advisor. Individuals can increase the rate of return in two ways. First, they can exert investment effort by reducing leisure time. Second, they can spend money in the first period at the expense of first period consumption.

In all three cases, individuals pay the financial advisor in the first period which will increase capital income in the second period. The individual budget constraint in period 1 and 2 are now, respectively

$$c_1 = Y - s - m - t,$$
$$c_2 = s + k(m, \cdot) - T,$$

where $k(\cdot)$ has the following properties[14]

$$k_m, k_{sm} > 0,\ k_{ma} \geq 0,\ k_{mm} \leq 0.$$

An increase in m leads to a higher rate of return, hence $k_{sm} > 0$. Buying financial information leads to an increase in capital income, but at a weakly decreasing rate. It is not quite clear whether k_{ma} should be positive or zero, while it seems obvious to rule out that more able investors receive less return from m.

I make no particular assumption on the sign of k_{Em} as some results below are sensitive to the sign. When individuals and their consultants somehow cooperate, then E and m are complements ($k_{Em} > 0$). It could also be the case that capital income depends on the total input of investment effort of individuals and their financial advisors, then E and m would be substitutes ($k_{Em} < 0$). Finally, it could well be that investment effort by individuals and financial advisors are completely independent such that $k_{Em} = 0$.

I will consider optimal allocations in a two type model, where individuals either differ in w or a. The setup is essentially the same as in section 4. The government observes Y, s and K.[15] All the comparative statics as well as the government's problem and the necessary conditions are presented in appendix D. As before, the no distortion at the top result holds. The question therefore boils down to the distortion faced by the less productive worker or less able investor type.

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 the results are case dependent.

Starting with case 1, it is shown in appendix D that mimickers have the same m as the type being mimicked. Therefore, intertemporal allocations should not be distorted and the Atkinson-Stiglitz theorem applies. This is because the intertemporal allocation does not depend on productivity, only income. This means

[14] Other properties are unchanged from those described in section 2.

[15] Expenditure on financial advisors is not tax deductible. A different approach might be to make the costs tax deductible such that the government observes $k(m, \cdot) - m$ instead of $k(m, \cdot)$. This approach would give mostly the same insights only that the analytics are more tedious.

that the simple fact that individuals have heterogeneous returns is not a sufficient argument for taxing capital. For optimal taxation, it matters why people have different rates of return. As in the Atkinson-Stiglitz model, labour income is distorted downwards in this model.

In case 2, a different result emerges. In the case where individuals differ only in terms of labour market productivity, the mimicker will not behave differently in terms of m. This means that capital income as well as wealth should be left undistorted. The reason is that capital income does not give any information on the underlying distribution of productivity. The intertemporal choices (s and m) solely depend on income and not productivity.

Turning to case 2, where individuals differ in terms of investment ability both savings and capital income should both be distorted downwards. The optimal intertemporal allocation is characterized by the following conditions

$$MRS_c^1 = 1 + k_s^1 - \frac{\gamma \hat{\psi}'_{21}}{n^1 \lambda_2} \left[MRS_c^1 - M\hat{R}S_c^{21} \right] < 1 + k_s^1 \tag{16a}$$

$$= k_m^1 - \frac{\gamma \hat{\psi}'_{21}}{n^1 \lambda_2} \left[MRS_c^1 - M\hat{R}S_c^{21} \frac{k_m^1}{\hat{k}_m^{21}} \right] < k_m^1. \tag{16b}$$

This means that the intertemporal allocation should be distorted downwards, as $MRS_c^i = 1 + k_s^i = k_m^i$ in the first best. Whether the total distortion on m should exceed the distortion on s depends on $\frac{k_m^1}{\hat{k}_m^{21}} \leq 1$.[16] If $\frac{k_m^1}{\hat{k}_m^{21}} < 1$, m faces a larger distortion than s, while the distortions are equal when $\frac{k_m^1}{\hat{k}_m^{21}} = 1$.

Mimickers mechanically spend less on financial advisors than the type being mimicked. Therefore, they have a higher return on m if $k_{mm} < 0$. In addition to this mimickers may also have a higher return on m due to the direct effects of a higher investment ability if $k_{ma} > 0$. This is the case when individual judgement is needed in choosing an advisor or different options provided by advisors.

Both s and m are downward distorted because more able investors have a different intertemporal allocation; they spend less on m (conditional on income). Therefore, they are more willing to increase s and m. This calls for a downward distortion on s and m. In addition to this, more able investors have, conditional on income, a greater return on their spendings on m if $\frac{dk_m}{da} > 0$. If this is the case then mimickers are more willing to increase m. This then calls for an even greater distortion on m compared to s.

How does this structure of distortions achieved by a tax system. I analyse that

[16] This is found by the following derivative: $\frac{dk_m}{da} = k_{mm}\frac{dm}{da} + k_{ma}$, where $\frac{dm}{da} < 0$, $k_{mm}, -k_{ma} \leq 0$. If either $k_{mm} < 0$ or $k_{ma} > 0$, then $\frac{dk_m}{da} > 0$ and then $\frac{k_m^1}{\hat{k}_m^{21}} < 1$.

by considering consumption in the second period in terms of a tax function,

$$c_2 = s + k(m, s, a) - T(s, k(m, s, a)). \tag{17}$$

It follows from the individual's necessary conditions[17] and (16b)-(16a) that $T_K > 0$, T_s has an ambiguous sign when $\frac{dk_m}{da} > 0$ and $T_s > 0$ when $\frac{dk_m}{da} = 0$. As long as $\frac{dk_m}{da}$ is not too large, then $T_s > 0$. This means that capital income and wealth may face a positive marginal tax rate.

Case 3 is the most general model where I have added the possibility to buy investment information to the model presented in section 4. The optimal allocation in this model take the same form as in (16a) and (16b), with the difference that the directions of the inequality may differ. The reason is that the comparative statics are less straightforward, largely because the sign of k_{Em} is ambiguous. First, it is not clear whether c_1 is higher for mimickers than type 1 individuals. Second, it is not clear whether k_m is lower for mimickers than type 1 individuals. This applies both to the model where individuals differ in w and a.

When individuals differ in terms of labour market productivity savings are distorted downwards, i.e. the inequality in (16a) holds. This is because MRS_c is decreasing in w. Individuals with higher productivity will, conditional on income, supply less labour and will therefore choose a higher E and a lower m to earn a given capital income. Regarding the distortion on the choice of m, it matters whether $k_m^1 \lesseqgtr \hat{k}_m^{21}$ as this will indicate the difference in the marginal willingness to increase m between mimickers and type 1 individuals. Because mimickers choose a higher E, the marginal return on m is higher for mimickers if E and m are complements ($k_{Em} > 0$), but if E and m are strong enough substitutes then $k_m^1 > \hat{k}_m^{21}$. As long as the $k_m^1 - \hat{k}_m^{21}$ is not too large, then m should also be distorted downwards, then inequality in (16b) also holds. Turning to the signs of the marginal tax rate by using the tax functions from (17), it is certain that $T_K > 0$ while the sign of T_s depends on $k_m^1 - \hat{k}_m^{21}$. If $k_m^1 - \hat{k}_m^{21}$ is not too large, then $T_s > 0$.

Turning to the model where individuals differ in terms of investment ability, the comparative statics is less straightforward. Mimickers will spend less on m as long as E and m are not too strong substitutes. The intuition is the following: Mimickers need a lower E in order to earn K^1, this reduces k_m (if E and m are substitutes) and then m has to increase in order to earn K^1. As long as this effect is not too strong, it will be counteracted by other effects and then mimickers will spend less on m. If this is the case, then mimickers have a lower MRS_c and then savings should be distorted downwards, i.e. the inequality in (16a) remains.

[17] The individual's necessary conditions imply that $MRS_c^i = k_m^i(1-T_K) = 1+k_s^i-(k_s^iT_K+T_s)$.

Regarding the optimal distortion on m, an additional ambiguity comes up because it is unclear whether mimickers have a higher or lower k_m unless $k_{Em} = 0$, then mimickers have a higher k_m, i.e. $\hat{k}_m^{21} > k_m^1$ (see further appendix D for discussion). As long as $k_m^1 - \hat{k}_m^{21}$ is not too large, then m will be distorted downwards. The results are summarized in proposition 4.

Proposition 4 *When individuals can spend money on financial advisors and the government observes savings, capital and labour income, the optimal tax system has the following case specific structure*

- *When there is heterogeneity in a or w and case 1 applies or heterogeneity in w and case 2 applies, there is no taxation of wealth nor capital income.*
- *When there is heterogeneity in a and case 2 or 3 applies or if there is heterogeneity in w and case 3 applies, savings and capital income should be distorted downwards (as long as E and m are not too strong substitutes).*

6.2 Domestic Credit Market

In the models I have considered so far, the domestic economy faces a market imperfection since there is no domestic credit market. Individuals with high rates of return should find it beneficial to borrow from individuals with lower rates of return. Both individuals could be made better off with such a transaction. Here I consider a model where in addition to the international investment market, there exists a domestic credit market. In the domestic credit market individuals can borrow and lend from each other at an interest rate r that is endogenously determined. I want to know whether the previous results change by adding the domestic credit market.

Individuals can borrow, or lend, in the first period and have to repay the loan in the second period as well as interest payments (or receive interest payments, in the case of a lender). The budget constraint of an individual is

$$c_1 = Y - s + b - t,$$
$$c_2 = s - b(1 + r) + k(E, s, a) - T.$$

where b is the amount borrowed (if $b > 0$) or loaned (if $b < 0$) and r is the interest rate that emerges on the domestic credit market, i.e. the equilibrium interest rate.

The government observes labour income and capital income and interest payments are tax deductible. This means that the capital income tax base is $K = k(E, s, a) - br$. This means that it is being assumed that the government does not know capital income from the international investment market separately from capital income from the domestic credit market. In reality, governments do have some knowledge on the nature of capital income. But if the government wants to tax different forms of capital income homogeneously to preserve neutrality, the assumed informational structure is reasonable.

Similarly to the setup in section 5.1, individuals choose E, s and b subject to the constraint $K = k(E, s, a) - br$. In appendix E, I set up the individual's problem and perform various comparative statics. I show that the optimal choice of b will depend on all the exogenous parameters, i.e. $b^i = b(Y^i, B_1^i, K^i, B_2^i, r, w^i, a^i)$.

As the domestic credit market only operates domestically, all aggregate lending has to equal aggregate borrowing. Therefore, the domestic credit market has the following equilibrium condition

$$\sum_i n^i b(Y^i, B_1^i, K^i, B_2^i, r, w^i, a^i) = 0. \tag{18}$$

The interest rate r, is the equilibrium interest rate. In other words, r is the rate that makes the supply equal the demand. In the absence of taxes, the domestic credit market will ensure that in equilibrium rates of return will be homogeneous, i.e. $r = k_s$.

In order to discuss the optimal tax system in this model, I need to know how changes in K and B_2 affect b. The effects of K and B_2 are twofold. First, there are direct effects, which are both positive (irrespective of whether b is positive or negative).[18] When K increases, individuals will mechanically save more in order for $k(E, s, a) - rb$ to equal K. This will shift consumption from the first period to the second period. Individuals will find it optimal to partly offset the increase in s by an increase in b.

When B_2 increases, individuals will have more second period consumption. As individuals have concave utility functions, they prefer to smooth their consumption stream and therefore they will increase b to shift consumption from the second period to the first period.

Second, there are indirect effects since changes in K and B_2 will affect r. This can be seen from the equilibrium condition (18). To do this it needs to acknowledged that there are borrowers and lenders in the economy which react differently to changes in in r. Changes in r will have income and substitution effects. To present the Slutsky equation, I define the expenditure function $x(B_1, K, V, r)$, which is the minimum level of B_2 required to attain a certain level of V. By the envelope theorem, the derivative of $x(B_1, K, V, r)$ w.r.t. r is (by implicitly

[18] This explanation is based on the individual problem where b and E are choice variables and s follows from the constraint $k(E, s, a) - br - K = 0$.

differentiating V)

$$\left.\frac{dx}{dr}\right|_{\overline{U}} = b\left[\frac{1}{k_s}\left(\frac{u'}{\psi'} - 1\right) + 1\right],$$

where the bracket is positive. Since there are borrowers and lenders, the sign of $\left.\frac{dx}{dr}\right|_{\overline{U}}$ will differ for borrowers and lenders, it will be negative for lenders ($b < 0$) and positive for borrowers ($b > 0$). The Slutsky equation for the effects of changes in r on b is

$$\frac{\partial b}{\partial r} = \left.\frac{\partial b}{\partial r}\right|_{\overline{U}} - \left.\frac{dx}{dr}\right|_{\overline{U}} \frac{\partial b}{\partial B_2},$$

where $\left.\frac{\partial b}{\partial r}\right|_{\overline{U}} < 0$ and $\frac{\partial b}{\partial B_2} > 0$ for lenders and borrowers. All the comparative statics are shown in appendix E. The Slutsky decomposition closely resembles the standard textbook model of borrowers and lenders (see e.g. Sandmo, 1985). The Slutsky equation takes a slightly different form from the standard model since $c_1 = B_1 - s(E, b, K, a, r) + b$ depends on r and $c_2 = B_2 + s(E, b, K, a, r) - b(1 + r)$ depends on r both directly and through $s(E, b, K, a, r)$.

A compensated increase in r will reduce b, i.e. $\left.\frac{\partial b}{\partial r}\right|_{\overline{U}} < 0$, both for lenders and borrowers. An increase in r makes borrowing more costly for borrowers and therefore they will reduce the compensated demand for b when r increases. An increase in r makes lending more profitable for lenders and therefore they will increase the compensated supply for b (which means that b decreases).

The direction of the income effects depend on whether individuals are borrowers or lenders. For borrowers, the income effects are negative as a rise in r will make them worse off. Thus, for a borrower it is clear that an increase in r implies reduced borrowing. This is shown graphically in figure 5 by the downward sloping demand curve.

For lenders, the income effects are positive and thus the relationship between b and r is ambiguous since substitution effects and income effects go in opposite directions. This means that the supply curve for lenders can be backward bending and in general, multiple equilibria cannot be ruled out. Here, I will only consider the case of a unique equilibrium, where the equilibrium occurs at the upward sloping part of the supply curve of b. This is the case depicted in figure 5.

Now I can analyse how changes in K or B_2 will affect r. For borrowers, an increase in K or B_2 will shift the demand curve upwards. A borrower will now want to borrow more for a given r. For lenders, a rise in K or B_2 will shift the demand curve downwards. Since they now want to lend less at a given r.

From figure 5, it can be seen that if K or B_2 increases for either the borrower or lender, the equilibrium r will increase.[19] This means that $\frac{\partial r}{\partial K^i} > 0$ and $\frac{\partial r}{\partial B_2^i} > 0$ for all types. This is an important fact for the results below.

[19] The formal analysis behind the comparative statics in figure 5 is done by totally differentiating condition (18) w.r.t. K and B_2:

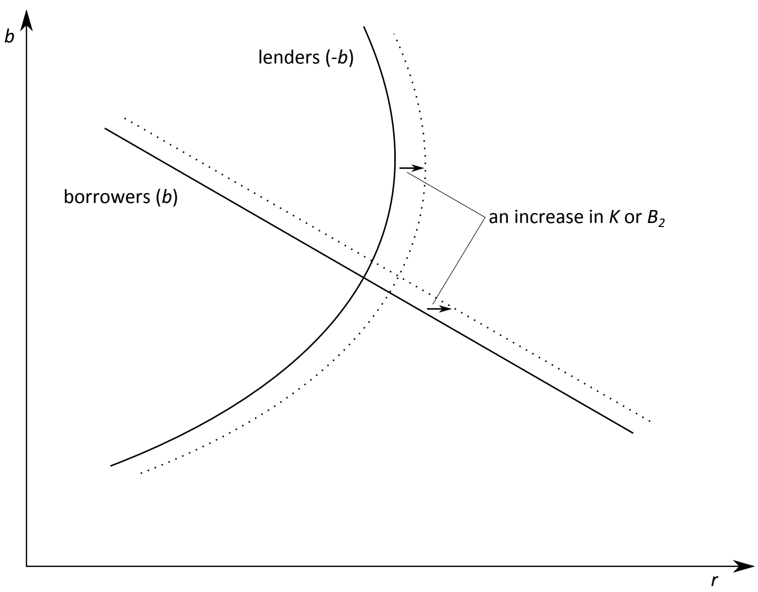

Figure 5: Effects of a rise in K or B_2 on r.

Having performed the positive analysis in this model, I turn to the normative analysis. In the first best, the same conclusions emerge as in section 3 and the domestic credit market is redundant since the government can do all redistribution between types directly. See appendix E for further discussion.

I consider a two type model where individuals differ either in w or a. The results presented below will hold for both models. In appendix E, I derive the necessary conditions for the government's problem. The optimal intertemporal allocation is

$$
\begin{aligned}
MRS_c^1 =& (1 + k_s^1) - \frac{\gamma \hat{\psi}_{21}}{\lambda_2 n^1} \left[(MRS_c^1 - \hat{MRS}_c^{21}) + \left(1 - \frac{k_s^1}{\hat{k}_s^{21}}\right) (\hat{MRS}_c^{21} - 1) \right] \qquad \text{(19a)} \\
& - \frac{\theta}{\lambda_2 n^1} \left[\frac{\partial r}{\partial K^1} k_s^1 + \frac{\partial r}{\partial B_2^1} (MRS_c^1 - 1) \right], \\
MRS_c^2 =& (1 + k_s^2) - \frac{\theta}{\lambda_2 n^2} \left[\frac{\partial r}{\partial K^2} k_s^2 + \frac{\partial r}{\partial B_2^2} (MRS_c^2 - 1) \right], \qquad \text{(19b)}
\end{aligned}
$$

where $\theta = \sum n^i b^i \psi_i' + \gamma(\psi_i' b^2 - \hat{\psi}_{21}' \hat{b}^{21})$, which can be either positive or negative. Both the brackets in (19a) are positive as well as the bracket in (19b) is positive. Since the sign of θ is ambiguous, it is not certain whether $MRS_c^i \lesseqgtr 1 + k_s^i$ for both types. Thus, the no distortion at the top result does not hold in this model.

$$
n^i \frac{\partial b^i}{\partial K^i} + \left[\sum n^i \frac{\partial b^i}{\partial r} \right] \frac{\partial r}{\partial K^i} = 0, \qquad n^i \frac{\partial b^i}{\partial B_2^i} + \left[\sum n^i \frac{\partial b^i}{\partial r} \right] \frac{\partial r}{\partial B_2^i} = 0,
$$

as long as $\sum n^i \frac{\partial b^i}{\partial r} < 0$, then $\frac{\partial r}{\partial K^i} > 0$ and $\frac{\partial r}{\partial B_2^i} > 0$ since $\frac{\partial b^i}{\partial K^i} > 0$ and $\frac{\partial b^i}{\partial B_2^i}$. This is the case depicted in figure 5.

The second term on the RHS in condition (19a) is an identical term as appears in condition (15a). In appendix E, I show that the term is positive leading to a downward distortion, just as in (15a). This means that the existence of a domestic credit market does not change the insights from section 5.

The bracket term $\frac{\partial r}{\partial K^i}k_s^i + \frac{\partial r}{\partial B_2^i}(MRS_c^i - 1)$ is positive, which indicates that an increase in MRS_c^1, will lead to an increase in r. In terms of taxes, this means that an increase in the marginal tax rate decreases r. Whether this is desirable or not depends on the sign of θ. A positive (negative) θ indicates that the rise in r is socially undesirable (desirable). θ consists of two terms. The first term is $\sum n^i b^i \psi_i'$, which is the direct welfare effect of a higher r. If the less skilled type is the borrower (if $b^1 > 0$), an increase in r is not desirable since it redistributes resources from the less skilled type (the borrower) to the more skilled type (the lender). If this is the case, this is an argument for a downward distortion on the intertemporal allocation. The opposite is true if the less skilled type is a lender. The second term is $\gamma(\psi_i' b^2 - \hat{\psi}_{21}' \hat{b}^{21})$ and indicates the effects of a rise in r on the incentive constraint. This term will be opposite to the first term if the mimicker has the same sign of b as the less skilled. If a rise in r will redistribute resources from the less skilled to the more skilled (which is undesirable), then it is likely that this will facilitate a redistribution from the mimicker to the more skilled type. This would relax the incentive constraint and therefore be desirable. This means that those effects tend to go in opposite directions. Therefore the sign of the third term in (19a) and the second term in (19b) have ambiguous signs irrespective of which is the borrower and lender. It should be expected that if individuals differ in terms of labour market productivity, then the less productive worker will be a borrower, then I would expect $\theta > 0$. On the other hand, if the individuals differ in terms of investment ability, it should be expected that the less able investor is a lender, then I would expect $\theta < 0$. The results are summarized in proposition 5.

Proposition 5 *When there is a domestic credit market, in addition to the international investment market, and the government observes capital and labour income, but not wealth, the optimal marginal tax rate consists of two terms. First, a positive term. Second, a term with an ambiguous sign that accounts for the welfare effects of taxes on r (the domestic interest rate). The term implies a positive marginal tax rate if a rise in r is socially undesirable, because a positive marginal tax rate will lead to a decrease in r.*

7 Conclusion

I have addressed 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. The analysis shows that the Atkinson-Stiglitz theorem that capital income should not be taxed does not hold when individuals exert investment effort. The application demonstrates that capital income should be taxed even if preferences are separable between leisure and consumption. Importantly, this result holds also when investment ability is homogeneous. If wealth is observable, it should be subsidized or taxed at the margin, depending on circumstances.

In the Atkinson-Stiglitz model, capital income does not reveal information about individual's underlying skill level. Individuals who have high capital income have so because they saved a lot. Therefore, the government should only tax labour income. In my model, this is not the case. Individuals may have high capital income because they save a lot, exert much investment effort or have high investment ability. In the Mirrlees model, labour income is distorted at the optimum because individuals with high labour market productivity are, conditional on income, at the margin more willing to work. This means that labour income will provide the government with information about individuals productivity. In my model the same holds for capital income, more skilled individuals are, conditional on income, at the margin more willing to save, exert investment effort and spend money on financial advisors. Therefore, I argue that capital income should be taxed for a similar reason labour is taxed in the Mirrlees model.

The results suggest that marginal tax rates for labour and capital income should not be equal but should be positively related to each other. The higher the marginal tax rate on labour income is, the higher the marginal tax rate on capital income should be.[20] Which marginal tax rate should be higher depends on the distribution of skills. If the inequality in labour market productivity is large, compared to the inequality in investment ability, the marginal tax rate on labour income should be higher and vice versa.

A distinct feature of the model is that a capital income tax and a wealth tax have different functions. Guvenen et al. (2017) is to the best of my knowledge the only other model to have such a feature.[21] In a model with perfect capital

[20] This is in line with the results by Banks and Diamond (2010: 550): "We lean towards relating marginal tax rates on capital and labour incomes to each other in some way (as in the US), as opposed to the Nordic dual income tax where there is a universal flat rate of tax on capital income."

[21] With the exception of taxing initial wealth in dynamic Ramsey models. In such models, it is optimal to tax initial wealth since this will not cause any behavioural effects if tax policy is time consistent.

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 all the models with heterogeneous labour market productivity and heterogeneous investment ability, capital income should be taxed while the sign of the marginal tax rate on wealth is case dependent.

This paper is mainly concerned with a two type model, where I have kept either labour market productivity or investment ability fixed. I only explore the full four type model in the model where the government does not observe wealth. Though this may seem somewhat limited the results apply in more general settings. All results also apply with a general distribution of types where labour market productivity and investment ability is perfectly correlated and when the correlation is sufficiently high such that productivity and ability is non-decreasing between adjacent types.

The key for my result that capital income should be taxed is the market imperfections embedded in the model. The capital market in the baseline model is imperfect as the returns are not equalized across individuals. It follows trivially that individuals would benefit from interpersonal lending, in which case returns would be equalized. The extension considered in section 6.3 therefore takes me somewhat away from the initial motivation of heterogeneous returns. Arguably, the reality lies somewhere in between the baseline model and the model with a domestic credit market. The remarkable result from section 6.3 is that even though there would exist a domestic credit market correcting for the imperfection stemming from the international investment market, there would still be a scope for taxing capital income. On the other hand, the sign of the tax is in general ambiguous. Interestingly, the insights that are obtained in section 5 are still valid if there is a domestic credit market, only that there will be additional effects on the domestic interest rate that need to be taken into account.

The paper most closely related to mine is Gahvari and Micheletto (2016). They point out that their result (of positive taxation of capital income) also holds when returns are an increasing function of effort. Hence, they state one of my main results (part of proposition 1). My contribution compared to their paper is to have a much more general setup.[22] In addition, I have a more thorough discussion of assumptions and results, whereas their paper deals first and foremost with a model where return on capital is increasing in labour income.

The current paper is a contribution to a growing literature on optimal taxation

[22]In my model people differ in their investment ability, the capital income tax is nonlinear instead of linear (which allows comparison of different marginal tax rates), I allow for the observability of wealth and I have two extensions.

of capital income. The literature has identified a number of conditions under which the Atkinson-Stiglitz theorem does not hold. A closely related model is presented in Christiansen and Tuomala (2008). In their model, individuals can, at some cost, legally shift income between tax bases. Reported capital income consists of true capital income and shifted labour income and therefore reflects the underlying skill distribution. In my model, people spend time to manage their portfolio by converting effort (or money on financial advisors) into higher capital income.

Another related strand of literature deals with human capital investment and optimal taxation. In this literature, individuals spend time and/or resources on education to increase their future productivity. In my model individuals spend time on increasing future capital income. The main difference is that human capital investment directly interacts with labour supply, which is not present in my model. Therefore, it may be optimal to subsidize education because it mitigates the discouraging effects of the labour income tax on human capital investment (Bovenberg and Jacobs, 2005). If human capital investment entails financial costs, taxing labour income distorts the choice between investing in human and financial capital in favour of the latter. Taxing capital income will then partly alleviate the distortion of the labour income tax on human capital investment (Jacobs and Bovenberg, 2010).

References

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

Bach, L., Calvet, L. E. and Paolo, S. (2015). "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." HEC Paris Research Paper No. FIN-2016-1126; Swedish House of Finance Research Paper No. 16-03.

Banks, J. and Diamond, P. (2010). "The Base for Direct Taxation." In *Dimensions of Tax Design: The Mirrlees Review*, ed. by the Institute for Fiscal Studies, 548-648. Oxford: Oxford University Press for the Institute for Fiscal Studies.

Bovenberg, L. and B. Jacobs (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics*, 89(11-12), 2005-2035.

Christiansen, V. (1984). Which Commodity Taxes should supplement the Income Tax? *Journal of Public Economics*, 24(2), 195-220.

Christiansen, V. and M. Tuomala (2008). On taxing capital income with income shifting. *International Tax and Public Finance*, 15(4), 527-545.

Cremer, H., Lozachmeur, J.-M. and Pestieau, P. (2004). Social Security, Retirement Age and Optimal Income Taxation. *Journal of Public Economics*, 88(11), 2259-2281.

Deuflhard, F., Georgarakos, D. and R. Inderst (2015). “Financial literacy and savings account returns.” Working Paper Series 1852, European Central Bank.

Diamond, P. and E. Saez (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives*, 25(4), 165-190.

Fagereng, A, Guiso, L., Malacrino, D. and L. Pistaferri (2016). "Heterogeneity and Persistence in Returns to Wealth," NBER Working Papers 22822, National Bureau of Economic Research, Inc.

Gahvari, F. and L. Micheletto (2016). Capital income taxation and the Atkinson–Stiglitz theorem. *Economics Letters*, 147, 86-89.

Gerritsen, A., Jacobs, B., Rusu, A., and K. Spiritus (2015). “Optimal taxation of capital income when capital returns are heterogeneous.” *Mimeo.*

Guiso, L. and T. Jappelli (2008). “Financial literacy and portfolio choice.” EUI Working papers no. ECO2008/31.

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo-Diaz, S. and D. Chen (2017). “Use It or Lose It: Efficiency Gains from Wealth Taxation.” *Mimeo.*

Hastings, J. S., Madrian, B. C., E. L. Skimmyhorn (2013). Financial Literacy, Financial Education, and Economic Outcomes. *Annual Review of Economics*, 5(1), 347-373.

Jacobs, B. and L. Bovenberg (2010). Human capital and optimal positive taxation of capital income. *International Tax and Public Finance*, 17(5), 451–478.

Lusardi, A. and O. S. Mitchell (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, 52(1), 5-44.

Mirrlees, J. (1971). An Exploration in the Theory of Optimal Income Taxation. *Review of Economic Studies*, 38, 175-208.

Piketty, T. (2014). *Capital in the Twenty-First Century.* Cambridge: Harvard University Press.

Piketty, T. and Saez, E. (2013). A Theory of Optimal Inheritance Taxation. *Econometrica*, 81(5), 1851-1886.

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

Sandmo, A. (1985). “The Effects of Taxation on Savings and Risk Taking.” In A. Auerbach, and M. Feldstein (eds.), *Handbook on Public Economics*, 265-311. North Holland: Elsevier Science Publishers.

Slemrod, J. and C. Gillitzer (2014). *Tax Systems*. Cambridge, MA: MIT Press.

Stern, N. (1982). Optimum Income 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), 213-40.

Stiglitz, J. E. (1985). Inequality and capital taxation, IMSSS Technical Report No. 457 (Stanford University).

Stiglitz, J. E. (1987). “Pareto Efficient and Optimal Taxation and the New New Welfare Economics.” In A. Auerbach, and M. Feldstein (eds.), *Handbook on Public Economics*, 991-1042. North Holland: Elsevier Science Publishers.

Stiglitz, J. E. (2015). In Praise of Frank Ramsey’s Contribution to the Theory of Taxation. *The Economic Journal*, 125(583), 235-268.

Stolper, O. and A. Walter (2017). Financial literacy, financial advice, and financial behavior. *Journal of Business Economics*, 87(5), 581-643.

Tuomala, M. (2016). *Optimal Redistributive Taxation*. Oxford: Oxford University Press.

Appendix

A Calculations for Section 4

Necessary Conditions In section 4 there are two different models. Both models have the same form and they are be solved in an identical manner. It is therefore sufficient to solve the problem once and then discuss the solutions for both models in separately. The Lagrangian for both models from section 4 is

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

From the condition $k(E^1, a^1, s^1) - k(\hat{E}^{21}, a^2, s^1) = 0$, it follows that $\frac{d\hat{E}^{21}}{dE^1} = \frac{k_E^1}{\hat{k}_E^{21}}$ and $\frac{d\hat{E}^{21}}{ds^1} = \frac{k_s^1 - \hat{k}_s^{21}}{\hat{k}_E^{21}}$. If $k(\cdot)$ is weakly separable such that $k(E, a, s,) = F(s, f(E, a))$ then $k_s^1 - \hat{k}_s^{21} = F_s(s^1, f(E^1, a^1)) - F_s(s^1, f(\hat{E}^{21}, a^2)) = 0$, where $f(E^1, a^1) = f(\hat{E}^{21}, a^2)$ since $F(s^1, f(E^1, a^1)) - F(s^1, f(\hat{E}^{21}, a^2))$.

The necessary conditions are

$$\begin{aligned}
(Y^1) &\qquad -n^1 v_1'/w^1 + \gamma \hat{v}_{21}'/w^2 + \lambda_1 n^1 = 0, \\
(Y^2) &\qquad -(n^2 + \gamma) v_2'/w^2 + \lambda_1 n^2 = 0, \\
(E^1) &\qquad -n^1 v_1' + \gamma \hat{v}_{21}' k_E^1/\hat{k}_E^{21} + \lambda_2 n^1 k_E^1 = 0, \\
(E^2) &\qquad -(n^2 + \gamma) v_2' + \lambda_2 n^2 k_E^2 = 0, \\
(s^1) &\qquad (n^1 - \gamma)(\psi_1' - u_1') + \lambda_2 n^1 k_s^1 = 0, \\
(s^2) &\qquad (n^2 + \gamma)(\psi_2' - u_2') + \lambda_2 n^2 k_s^2 = 0, \\
(B_1^1) &\qquad (n^1 - \gamma) u_1' - \lambda_1 n^1 = 0, \\
(B_1^2) &\qquad (n^2 + \gamma) u_2' - \lambda_1 n^2 = 0, \\
(B_2^1) &\qquad (n^1 - \gamma) \psi_1' - \lambda_2 n^1 = 0, \\
(B_2^2) &\qquad (n^2 + \gamma) \psi_2' - \lambda_2 n^2 = 0,
\end{aligned}$$

The optimal allocation in (6b), (6c) and (6e) follow immediately. 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 them together. Next, I multiply both sides with $(\gamma\psi_1' + \lambda_2 n^1)/\gamma\psi_1'$ and after simple manipulations, I get (6a). Equation (6d) is found by very similar algebra steps. As discussed in the beginning of section 4, mimickers will have more leisure than type 1 individual in both models, this means that $v_1' > \hat{v}_{21}'$. In the model where individuals differ in a, $k_E^1(E^1, a^1, s^1) < \hat{k}_E^{21}(\hat{E}^{21}, a^2, s^1)$. These two facts imply that $MRS_K^1 = \frac{v_1'}{\psi_1' k_E^1} > \hat{MRS}_K^{21} = \frac{\hat{v}_{21}'}{\hat{\psi}_{21}' \hat{k}_E^{21}}$. It also follows that $MRS_Y^1 = \frac{v_1'}{u_1' w^1} > \hat{MRS}_Y^{21} = \frac{\hat{v}_{21}'}{\hat{u}_{21}' w^2}$ in both models since $w^2 \geq w^1$.

Difference in Distortion between Y and K Noting that $MRS_K = MRS_Y \frac{w(1+k_s)}{k_E}$, it follows that finding out whether t' or T_K is larger crucially depends on $w(1+k_s) \lesseqgtr k_E$, i.e. in which direction the the marginal return to L and E is distorted. This turns out be different for the models presented in section 4.1. and 4.2. 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 10: $MRS_Y = 1 - t'$ and $MRS_K = 1 - T_K$, it follows that

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

B Calculations for Section 5

Sign of s_{Ea} Given the properties of $k(E, a, s)$ the sign of $s_{Ea} = (k_E dk_s/da - k_s dk_E/da)/k_s^2$ is ambiguous. If the following form of weak separability is satisfied

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

then $s_{Ea} > 0$. This will be assumed to be the case from now on.[23] It should be emphasized that the sign of s_{Ea} is purely a matter of the capital income function $k(E, a, s)$, i.e. it only depends on the economy's technology and not on preferences.

Deriving (13) To derive the inequality in (13), I differentiate (9) to determine E_a and use (7) to determine s_E and s_a. By some manipulation the inequality follows

$$\frac{ds}{da} U_{EE} = -U_{Ea} s_E + s_a U_{EE}$$

[23] This can be shown as follows

$$s_{Ea} = \left[k_E \frac{dk_s}{da} - k_s \frac{dk_E}{da} \right] / k_s^2 = \frac{k_E}{k_s} \left[\frac{k_{sa}}{k_s} - \frac{k_{Ea}}{k_E} \right] + s_a \frac{k_{ss} k_E - k_{Es} k_s}{k_s^2} > 0$$

where under the form (20): $\frac{k_{sa}}{k_s} - \frac{k_{Ea}}{k_E} = \frac{F_{12} f_E}{F_1 f_E} - \frac{F_{12} f_s}{F_1 f_s} = 0$, where F_1 is the partial derivative of F w.r.t. to it's 1st argument and F_{12} is the cross derivative.

$$= \frac{u' - \psi'}{k_s^2} s_a \left[k_{EE} k_s - k_{sE} k_E\right] + s_a v'' > 0 \implies \frac{ds}{da} < 0.$$

Necessary Conditions The Lagrangian for (14) 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 (15a) 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 with $(\gamma \hat{\psi}_{21}' + \lambda_2 n^1)/\lambda_2 n^1$ 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 (15a).

In order to show the inequality in (15a), I first need to show that $\hat{k}_s^{21} > k_s^1$, 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, $E(\cdot)$ is the individual's optimal choice, which is analyses by differentiating (9) and $s(\cdot)$ (which follows from the constraint (7)),

$$\begin{aligned}
\frac{dk_s}{da} =& k_{sE} E_a + k_{ss}(s_E E_a + s_a) + k_{sa}, \\
\frac{dk_s}{da} U_{EE} =& - k_{sE} U_{Ea} - k_{ss} s_E U_{Ea} + k_{ss} s_a U_{EE} + k_{sa} U_{EE} \\
=& \left[\frac{k_{sE}(u' - \psi')}{k_s^2} + \frac{s_E(u'' + \psi'')}{k_s}\right] \left[k_{sE} k_a - k_{sa} k_E\right] + \frac{u' - \psi'}{k_s^2} \left[k_{ss} s_a k_{EE} k_s \right. \\
& \left. - k_{sa} k_{ss} s_E k_E - k_{sa} k_{sE} k_E + k_{sa} k_{EE} k_s\right] + v'' \left[k_{ss} s_a + k_{sa}\right].
\end{aligned}$$

In the final expression the second and third term are negative and $k_{sE}(u' - \psi')/k_s^2 + s_E(u'' + \psi'')/k_s > 0$ but $k_{sE} k_a - k_{sa} k_E$ has an ambiguous sign. Assuming a stronger separability $k(E, s, a)$ such that $k(E, s, a) = f^1(E) f^2(s) g(a)$, then $k_{sE} k_a - k_{sa} k_E = 0$ and then $\frac{dk_s}{da} U_{EE} < 0$, which implies that $\frac{dk_s}{da} > 0$. It should be noted that the functional form for k is only a sufficient condition, as long as $k_{sE}k_a - k_{sa}k_E$ does not become too positive, I will have $\frac{dk_s}{da} > 0$. This shows that $\hat{k}_s^{21} > k_s^1$.

From the individuals necessary condition (9), it follows that $\psi' - u' = -v'E_s < 0$. This shows that $\left(1 - \frac{k_s^1}{\hat{k}_s^{21}}\right)\left(1 - \frac{\hat{u}'_{21}}{\hat{\psi}'_{21}}\right) < 0$. 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 $\frac{u'_1}{\psi'_1} > \frac{\hat{u}'_{21}}{\hat{\psi}'_{21}}$. Therefore, the bracket in (15a) is positive. This proves the inequality in (15a).

Showing that the inequality in (15a) also holds for the model where individuals differ in w, can be proven analogously, only that it is simpler to show that $\hat{k}_s^{21} > k_s^1$ since

$$\frac{dk_s}{dw} = k_{sE}E_w + k_{ss}s_E E_w > 0$$

Here, $E_w > 0$ follows from (9) and $s_E < 0$ from (7).

Condition (15b) 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 (15b).

In order to derive the inequality in (15b) for both the model with heterogeneous a and heterogeneous w, I note that $s^1 > \hat{s}^{21}$ and $E^1 + L^1 > \hat{E}^{21} + \hat{L}^{21}$, hence $u'_1 > \hat{u}'_{21}$ and $v'_1 > \hat{v}'_{21}$. Therefore, the sign of $MRS_Y^1 - M\hat{R}S_Y^{21}$ is ambiguous and therefore I cannot sign the direction of the inequality in (15b).

Difference in Distortion between Y and K Individuals receive labour income and capital income in the second period. As in the comparative statics performed in section 5.1, individuals face the problem (8). All the comparative statics are analogous in this model and no need to repeat them here. The government's budget constraint is now

$$\sum_i n^i(e - B_1^i) \geq g_1,$$
$$\sum_i n^i(Y^i + K^i - B_2^i) \geq g_2.$$

The governments optimal allocation in this model is (performing manipulations similar to the ones above)

$$MRS_Y^1 = 1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[MRS_Y^1 - M\hat{R}S_Y^{21}\right] = 1 - T_Y,$$
$$MRS_K^1 = 1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[MRS_K^1 - M\hat{R}S_K^{21}\right] = 1 - T_K,$$

where $MRS_Y = \frac{v'}{\psi'}\frac{1}{w}$ takes a slightly different form. Noting that $MRS_K = MRS_Y \frac{w}{k_E}$, it follows that finding out whether T_Y or T_K is larger crucially depends on $w \lesseqgtr k_E$, i.e. in which direction the the marginal return to L and E is distorted. This turns out be different for the models presented in section 5.2. and 5.3. Using the necessary conditions for Y^1 and K^1, I can write the following expression

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

In the model from section 5.2, I have $\frac{w^1}{w^2} < 1$ and $\frac{k_E^1}{\hat{k}_E^{21}} > 1$ since $\frac{dk_E}{dw} = k_{EE}E_w + k_{Es}s_E E_w < 0$, hence $w^1 - k_E^1 > 0$, i.e. the government distort the marginal return for L and E in favour of L. In the model from section 5.3., I have $\frac{w^1}{w^2} = 1$ and $\frac{k_E^1}{\hat{k}_E^{21}} < 1$ since

$$\begin{aligned}
\frac{dk_E}{da} =& k_{EE}E_a + k_{Es}(s_E E_a + s_a) + k_{Ea}, \\
\frac{dk_E}{da} U_{EE} =& - k_{EE}U_{Ea} - k_{Es}s_E U_{Ea} + k_{Es}s_a U_{EE} + k_{Ea}U_{EE} \\
=& (u' - \psi')[s_a k_{EE} k_{ss} k_E - k_{ss} s_E k_E k_{Ea} - k_{Es} k_E k_{Ea} + k_{Ea} k_{EE} k_s] \\
& + (u'' + \psi'')[s_E^2 k_{Ea} - s_E s_a k_{EE}] < 0, \implies \frac{dk_E}{da} > 0.
\end{aligned}$$

Hence, $w^1 - k_E^1 < 0$, i.e. the government distort the marginal return for L and E in favour of E.

Du to the necessary individual's conditions: $MRS_Y = 1 - T_Y$ and $MRS_K = 1 - T_K$, it follows that

$$\begin{aligned}
w^2 > w^2, a^2 = a^1 : w^1 > k_E^1 \;\Rightarrow\; MRS_K^1 < MRS_Y^1 \;\Rightarrow\; T_Y > T_K, \\
w^2 = w^2, a^2 > a^1 : w^1 < k_E^1 \;\Rightarrow\; MRS_K^1 > MRS_Y^1 \;\Rightarrow\; T_Y < T_K.
\end{aligned}$$

C Calculations for Section 5.3

The derivatives of MRS_Y w.r.t. w and a are

$$\begin{aligned}
\frac{\partial MRS_Y}{\partial w} =& \frac{v''u'w\left[\frac{Y}{w^2} - \frac{\partial E}{\partial w}\right] - v'u''w\left[u' - wu''\frac{ds}{dw}\right]}{(u''w)^2} \gtreqless 0, \\
\frac{\partial MRS_Y}{\partial a} =& \frac{v''u'w\left[-\frac{\partial E}{\partial a}\right] + v'u''w\left[\frac{ds}{da}\right]}{(u''w)^2} \gtreqless 0,
\end{aligned}$$

where $\frac{Y}{w^2} - \frac{\partial E}{\partial w}, -\frac{\partial E}{\partial a} > 0$ and $\frac{ds}{dw}, \frac{ds}{da} < 0$.

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

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

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 subsidized).

To simplify, I will assume that the incentive constraint on type 4 mimicking type 1 to be slack (this will not affect the main qualitative results). The necessary conditions for Lagrangian in formula (21) are

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

Manipulating the necessary conditions as in appendix A and B leads to the following optimal intertemporal allocations

$$MRS_c^1 =(1 + k_s^1) - \frac{\gamma^{21}\hat{\psi}_{21}'}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{21}) + (1 - k_s^1/\hat{k}_s^{21})(M\hat{R}S_c^{21} - 1)\right]$$

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

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. There will be 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 6.1

Case 1: $k(s, m)$

When the government observes Y, s and $K = k(s, m)$, individuals maximize utility w.r.t. m subject to the constraint $k(s, m) = K$, this implicitly defines $m(s, K)$. This means that the expenditure on m simply depends on the budget chosen and not on the type. That is, mimickers will have the same consumption stream as the type being mimicked.

Case 2: $k(m, s, a)$

When the government observes Y, s and $K = k(s, m, a)$, individuals maximize utility w.r.t. m subject to the constraint $k(s, m, a) = K$. This means that individuals simply choose the m that will give them the K reported by the government. By the constraint $k(s, m, a) = K$, I implicitly define $m(s, K, a)$, with

$$
\begin{aligned}
m_w &= 0,\\
m_a &= -\frac{k_a}{k_m} < 0.
\end{aligned}
$$

This means that mimickers differing in therms of w will have the same consumption stream as the type mimicked, while a mimicker with a higher investment ability will spend less on m and therefore have a higher first period consumption.

I solve the government's problem similarly as is done for the model in section 4. Instead of the government choosing K, they choose an m that corresponds to a certain value of K. As the government does not observe m, mimickers have a lower m, this follows from the condition $k(s^1, m^1, a^1) = k(s^1, \hat{m}^{21}, a^2)$. From this condition it follows that $\frac{d\hat{m}^{21}}{dm^1} = \frac{k_m^1}{\hat{k}_m^{21}} \leq 1$, since $\frac{dk_m}{da} = k_{mm}m_a + k_{ma} \geq 0$. Also, $\frac{d\hat{m}^{21}}{ds^1} = \frac{k_s^1 - \hat{k}_s^{21}}{\hat{k}_m^{21}} = 0$, assuming that k is weakly separable such that $k = F(s, f(a, m))$, then $k_s^1 = \hat{k}_s^{21}$.

The Lagrangian for this problem is

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

where $U^i = u(B_1^i - s^i - m^i) + \psi(B_2^i + s^i) + v(1 - Y^i/w)$ and $\hat{U}^{21} = u(B_1^1 - s^1 - \hat{m}^{21}) + \psi(B_2^1 + s^1) + v(1 - Y^1/w)$. The necessary conditions for type 1 are

$$\begin{aligned}(m^1) \qquad & -n^1 u_1' + \gamma \hat{u}_{21}' \frac{k_m^1}{\hat{k}_m^{21}} + \lambda_2 n^1 k_m^1 = 0, \\ (s^1) \qquad & -n^1 (u_1' - \psi_1') + \gamma(\hat{u}_{21}' - \psi_1') + \lambda_2 n^1 k_s^1 = 0, \\ (B_2^1) \qquad & (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0.\end{aligned}$$

Manipulation similar to those performed in appendix A will lead to (16a) and (16b).

Case 3: $k(E, m, s, a)$

The government observes Y^i, s^i and K^i, and offers bundles in terms of these variables for both types of individuals. I follow a similar procedure as in section 5. Individuals choose E and m, facing the constraint $k(E^i, s^i, m^i, a^i) = K^i$. This implicitly defines $E(K, s, m)$. Deriving partial derivatives is done by differentiating $k(E(K, s, m), s, m) - K = 0$. The individual's problem and necessary condition are, respectively

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

where $E_m = -k_m/k_E < 0$. As before: $B_1 = Y - t$ and $B_2 = K - T$. The problem is comprised by finding the optimal mix between m and E such that $k(E, s, m) - K = 0$.

Differentiating $U_m = 0$ gives the following comparative static results

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

where $E_a = -k_a/k_E < 0$, $E_{mm} = [(k_{EE}E_m + k_{Em})k_m - (k_{mE}E_m + k_{mm})k_E]/k_E^2$. Assuming that the second order conditions are satisfied, an increase in w leads to a reduction in m and an increase in E. This means that mimickers differing in terms of w enjoy more leisure and more c_1. The mechanical effect of mimicking is that L is lower than 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 $\frac{dL}{dw}$.

Performing the same comparative statics for an increase in a gives

$$\frac{dm}{da} = \frac{v'E_{ma} - v''E_mE_a}{U_{mm}} < 0, \quad \text{if } k_{mE} \geq 0, \text{ or not too negative,}$$

where $E_{ma} = [(k_{EE}E_a + k_{Ea})k_m - (k_{mE}E_a + k_{ma})k_E]/k_E^2 = [k_{mE}k_a + k_{EE}k_mE_a]/k_E^2$, assuming that $k(\cdot)$ is weakly separable such that $k(\cdot) = f_1(s)f_2(a)g(E, m)$. In general, the sign of E_{ma} is ambiguous as it depends on the sign of k_{mE}. If $k_{mE} \geq 0$ then $E_{ma} > 0$ but if $k_{mE} < 0$ is sufficiently negative then $E_{ma} < 0$. As long as $E_{ma} > 0$ or not too negative, then $\frac{dm}{da} < 0$. The intuition is the following: An increase in a leads to a mechanical reduction in investment effort due to the constraint $k(E, s, m, a) - K = 0$ and therefore leisure time increases. This leads to an increase in k_m (if $k_{Em} < 0$) and therefore spendings on m have to increase so that earned capital income remains K. This is counteracted by two other effects: First, k will increase (if $k_{EE} < 0$) and therefore spendings on m have to decrease so that earned capital income remains K. Second, due to the concavity in the utility function, it is optimal to partly use this utility gain caused by more leisure time by reducing m.

Performing the same comparative statics for the effects of an increase in a on E give

$$\frac{dE}{da} = E_m\frac{dm}{da} + E_a = \frac{u''E_a + v'[E_{ma}E_m - E_{mm}E_a]}{U_{mm}}$$
$$= \frac{u''E_a}{U_{mm}} + \frac{v'E_a}{U_{mm}k_E}[k_{mm} + k_{mE}E_a] < 0, \quad \text{if } k_{mE} \geq 0.$$

The first term is unambiguously negative and the second term is unambiguously non-positive if $k_{mE} \geq 0$. This means that as long as E and m are not too strong substitutes (i.e. k_{Em} is not too negative), then $\frac{dE}{da} < 0$. Taken together, as long as k_{Em} is not too negative, mimickers differing in terms of a will have a lower m and E. The intuition is the following: The mechanical effect of mimicking is that the individual needs less E to produce K. 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 the pure mechanical effect.

I define the indirect utility function by V, it is derived plugging the optimal m into U. The derivatives are

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

The Lagrangian for the government's problem is

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

The necessary conditions for type 1 are

$$\begin{aligned}&(K^1) \qquad -\frac{n^1 u_1'}{k_m^1} + \frac{\gamma \hat{u}_{21}'}{\hat{k}_m^{21}} + \lambda_2 n^1 = 0, \\ &(s^1) \qquad -n^1 u_1'\left(1 - \frac{k_s^1}{k_m^1}\right) + \psi_1' + \gamma\hat{u}_{21}'\left(1 - \frac{\hat{k}_s^{21}}{\hat{k}_m^{21}}\right) - \gamma\psi_1' = 0, \\ &(B_2^1) \qquad (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0.\end{aligned}$$

Performing similar derivations as in appendix A on the necessary conditions for K^1 and s^1, I get the condition (16a). Solving for $n^1 u_1'$ in the necessary condition for K^1 and making use of necessary conditions for s^1 and B_2 yields

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

Solving for $n^1\psi_1'$ from the necessary condition for B_2 and diving with the above condition and performing similar derivations as in appendix A gives condition (16b) if $\hat{k}_s^{21} - k_s^1 = 0$, which will be the case if $k = f_1(s)f_2(a)g(E, m)$ is weakly separable,

$$\begin{aligned}\frac{dk_s}{dw} =& k_{sE}\frac{dE}{dw} + k_{sm}\frac{dm}{dw}\\ =& \frac{dm}{dw}\left[k_{sm} - \frac{k_{sE}k_m}{k_E}\right] = 0,\\ \frac{dk_s}{da} =& k_{sE}\frac{dE}{da} + k_{sm}\frac{dm}{da} + k_{sa}\\ =& \frac{dm}{da}\left[k_{sm} - \frac{k_{sE}k_m}{k_E}\right] + \left[k_{sa} - \frac{k_{sE}k_a}{k_E}\right] = 0,\end{aligned}$$

this shows that $\hat{k}_s^{21} - k_s^1 = 0$ and therefore the optimal allocation is described by (16a).

Finally, I need to know whether $\frac{k_m^1}{\hat{k}_m^{21}} \lesseqgtr 1$. I analyse that by the following derivatives

$$\begin{aligned}\frac{dk_m}{dw} &= k_{mE}\frac{dE}{dw} + k_{mm}\frac{dm}{dw} > 0, \quad \text{if } k_{mE} \geq 0, \text{ or not too negative},\\ \frac{dk_m}{da} &= k_{mE}\frac{dE}{da} + k_{mm}\frac{dm}{da} + k_{ma} > 0, \quad \text{if } k_{mE} = 0.\end{aligned}$$

This shows that the sign of $\frac{k_m^1}{\hat{k}_m^{21}} - 1$ depends on the sign of k_{mE}.

An exogenous increase in a will mechanically increase k_m(since $k_{ma} > 0$). If $k_{mE} \geq 0$, or not too negative, an increase in a will lead to a reduction in m, which leads to an increase in k_m (since $k_{mm} \leq 0$). If $k_{mE} \geq 0$, or not too negative, an increase in a will lead to a reduction in E. How this then affects k_m depends on the sign of k_{mE}.

E Calculations for Section 6.3

First best

The government's problem is

$$\begin{aligned}\max_{\{Y^i, B_1^i, E^i, s^i, b^i, B_2^i\}} \quad & \sum_i n^i V(Y^i, B_1^i, E^i, s^i, b^i, B_2^i, a^i, w^i)\\ \text{subject to} \quad & \sum_i n^i (Y^i + b^i - B_1^i) \geq g_1,\\ & \sum_i n^i (k(E^i, s^i, a^i) - b^i(1+r) - B_2^i) \geq g_2,\\ & \sum_i n^i b^i \leq 0.\end{aligned}$$

Noting that the constraint $\sum_i n^i b^i \leq 0$ will always be binding, b^i can simply be eliminated from the government's budget constraint in both periods.

The necessary conditions will be the same as in the third section in addition to the necessary condition for b^i, which implies that $MRS_c^i = 1 + r$. The necessary condition for s^i implies that $MRS_c^i = 1 + k_s^i$. This means that the government will have $r = k_s^i$. In other words, the interest rate on the domestic would be equal to the equilibrium rate of return. As the government controls s^i for all types, they are no better off having the control variable b^i in addition.

Second best

Government's necessary Conditions The Lagrangian to the government's problem is identical to the one in appendix B, only that now individuals can borrow and lend. Making use of the envelope theorem, the necessary conditions for K and B_2 are

$$
\begin{aligned}
(K^1) \qquad & -n^1(u_1' - \psi_1')/k_s^1 + \gamma(\hat{u}_{21}' - \hat{\psi}_{21}')/\hat{k}_s^{21} \\
& -\frac{\partial r}{\partial K^1}[n^1\psi_1' b^1 - \gamma\hat{\psi}_{21}'\hat{b}^{21} + (n^2+\gamma)\psi_2' b^2] + \lambda_2 n^1 = 0, \\
(K^2) \qquad & -(n^2+\gamma)(u_2' - \psi_2')/k_s^2 \\
& -\frac{\partial r}{\partial K^2}[n^1\psi_1' b^1 - \gamma\hat{\psi}_{21}'\hat{b}^{21} + (n^2+\gamma)\psi_2' b^2] + \lambda_2 n^2 = 0, \\
(B_2^1) \qquad & n^1\psi_1' - \gamma\hat{\psi}_{21}' - \frac{\partial r}{\partial B_2^1}[n^1\psi_1' b^1 - \gamma\hat{\psi}_{21}'\hat{b}^{21} + (n^2+\gamma)\psi_2' b^2] - \lambda_2 n^1 = 0, \\
(B_2^2) \qquad & (n^2+\gamma)\psi_2' - \frac{\partial r}{\partial B_2^2}[n^1\psi_1' b^1 - \gamma\hat{\psi}_{21}'\hat{b}^{21} + (n^2+\gamma)\psi_2' b^2] - \lambda_2 n^2 = 0.
\end{aligned}
$$

Performing manipulations similar to those in appendix B, will lead to equations (19a) and (19b).

Demand for b The individual problem can be presented similar to (9)

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

where $s(E,b,K,a,r)$ is implicitly defined by the constraint $k(E,s,a) - rb = K$. The necessary conditions are

$$
\begin{aligned}
U_E &= -u's_E + \psi's_E - v' = 0, \\
U_b &= u'(1 - s_b) - \psi'(1 + r - s_b) = 0,
\end{aligned}
$$

where $s_E = -k_E/k_s < 0$ and $s_b = r/k_s > 0$. The necessary conditions imply that $1 + r - s_b > 1 - s_b > 0$, and therefore $r < k_s$. These conditions define the optimal

b and E as a function of all the exogenous variables, $b(Y, B_1, K, B_2, r; w, a)$ and $E(Y, B_1, K, B_2, r; w, a)$. Below, I analyse how b responses to changes in the various exogenous variables.

First, I look at how b changes with r. I set $b(r)$ and $E(r)$ and differentiate the individual's necessary conditions and get the following system of equations

$$\begin{bmatrix} U_{EE} & U_{Eb} \\ U_{bE} & U_{bb} \end{bmatrix} \begin{bmatrix} E_r \\ b_r \end{bmatrix} = \begin{bmatrix} -U_{Er} \\ -U_{br} \end{bmatrix}.$$

The Hessian matrix is denoted by H. The following partial derivatives are used below

$$\begin{aligned}
&U_{EE} = (u'' + \psi'')s_E^2 - (u' - \psi')s_{EE} + v'' < 0, \\
&U_{bb} = u''(1 - s_b)^2 + \psi''(1 + r - s_b)^2 < 0, \\
&U_{bE} = -u''(1 - s_b)s_E - \psi''(1 + r - s_b)s_E - (u' - \psi')s_{bE}, \\
&U_{Er} = (u'' + \psi'')s_E s_r - \psi'' s_E b, \\
&U_{br} = -u''(1 - s_b)s_r - \psi''(1 + r - s_b)s_r - \psi' + \psi''(1 + r - s_b)b - (u' - \psi')s_{br}, \\
&U_{EB_2} = \psi'' s_E > 0, \\
&U_{bB_2} = -\psi''(1 + r - s_b) > 0, \\
&U_{EB_1} = -u'' s_E < 0, \\
&U_{bB_1} = u''(1 - s_b) < 0, \\
&U_{EK} = (u'' + \psi'')s_E s_K - (u' - \psi')s_{EK} > 0, \\
&U_{bK} = -u''(1 - s_b)s_K - \psi''(1 + r - s_b)s_K > 0,
\end{aligned}$$

where $s_{bE} = -r(k_{sE} + k_{ss}s_E)/k_s^2 < 0$, $s_r = b/k_s$, $s_{br} = 1/k_s$ and $s_{EK} = (k_E k_{ss} s_K - k_s k_{Es} s_K)/k_{ss}^2 < 0$.

Using Cramer's rule, I can solve for b_r

$$\begin{aligned}
b_r\,|H| =& U_{bE}U_{Er} - U_{EE}U_{br} \\
=& \psi''(u' - \psi')s_{bE}s_E b - u''(u' - \psi')[s_{bE}s_E + (1 - s_b)s_{EE}]s_r \\
&- \psi''(u' - \psi')[s_{bE}s_E + (1 + r - s_b)s_{EE}]s_r + v''u''(1 - s_b)s_r \\
&+ v''\psi''(1 + r - s_b)s_r + [\psi' + (u' - \psi')s_{br}]U_{EE} - u''\psi'' r s_E^2 b \\
&+ \psi''(u' - \psi')(1 + r - s_b)s_{EE}b - v''\psi''(1 + r - s_b)b,
\end{aligned}$$

where $|H| > 0$, which follows from the second order condition, which is assumed to hold. As there are both income and substitution effects, I cannot sign b_r in general. To look only at the compensated effects of an increase in r, I need to consider the effects of a change in B_1 and B_2

$$
\begin{aligned}
b_{B_1}|H| &= U_{bE}U_{EB_1} - U_{EE}U_{bB_1} \\
&= u''\psi''s_E^2 r + u''(u'-\psi')[s_E s_{bE} + (1-s_b)s_{EE}] - v''u''(1-s_b) < 0, \\
b_{B_2}|H| &= U_{bE}U_{EB_2} - U_{EE}U_{bB_2} \\
&= u''\psi''s_E^2 r - \psi''(u'-\psi')[s_{bE}s_E + s_{EE}(1+r-s_b)] + \psi''v''(1+r-s_b) > 0,
\end{aligned}
$$

the sign of b_{B_1} follows from the fact that $b_{B_2} = b_{B_1} \left.\frac{dB_1}{dB_2}\right|_{\overline{U}} = -b_{B_1}\frac{\psi'}{u'}$. To sign the compensated effect of r, I note that $\frac{\partial b}{\partial B_2}\left.\frac{dB_2}{dB_1}\right|_{\overline{U}} = -\frac{\partial b}{\partial B_1}$ and that

$$
\left.\frac{dx}{dr}\right|_{\overline{U}} b_{B_2} = -s_r b_{B_2}\left.\frac{dB_2}{dB_1}\right|_{\overline{U}} + (b-s_r)b_{B_2} = s_r b_{B_1} + (b-s_r)b_{B_2}.
$$

Now it can shown that a compensated increase in r will reduce b

$$
\begin{aligned}
b_r^c =& [b_r + x_r^c b_{B_2}] \\
=& [b_r + s_r b_{B_1} + (b-s_r)b_{B_2}] \\
=& [\psi' + (u'-\psi')s_{br}]U_{EE}|H|^{-1} < 0.
\end{aligned}
$$

Next, I show the effects of a change in K on b

$$
\begin{aligned}
b_K|H| =& U_{EK}U_{bE} - U_{bK}U_{EE} \\
=& [u''(u'-\psi') + \psi''(u'-\psi')(1+r)][k_{EE}k_s - k_{sE}k_E]s_K/k_s^2 \\
&+ [u'' + \psi''(1+r)]v''s_K > 0.
\end{aligned}
$$

Behaviour of Mimickers I follow the same procedure as above to show the effects of small changes in w and a on b and E. First, I note the following partial derivatives

$$
\begin{aligned}
U_{Ew} &= -v''Yw^{-2} > 0, \\
U_{bw} &= 0, \\
U_{Ea} &= (u''+\psi'')s_E s_a - (u'-\psi')s_{Ea} < 0, \\
U_{ba} &= -u''(1-s_b)s_a - \psi''(1+r-s_b)s_a - (u'-\psi')s_{ba},
\end{aligned}
$$

where $s_{ba} = -r(k_{ss}s_a + k_{sa})/k_s^2 < 0$. The effects of an increase in w on b and E are ambiguous and positive, respectively

$$
\begin{aligned}
b_w|H| &= U_{bE}U_{Ew}, \\
E_w|H| &= -U_{Ew}U_{bb} > 0.
\end{aligned}
$$

It follows that mimickers will also save less and therefore they will have a larger (smaller) first (second) period consumption, compared to the less able worker.

$$
\begin{aligned}
\frac{ds}{dw} =& s_E E_w < 0, \\
\left(\frac{ds}{dw} - b_w\right)|H| =& - s_E U_{Ew} U_{bb} - U_{bE} U_{Ew} \\
=& - U_{Ew}[u''(1-s_b)s_E(1-s_b-s_E) \\
& + \psi''(1+r-s_b)s_E(1+r-s_b-s_E) - (u'-\psi')s_{bE}s_E] < 0.
\end{aligned}
$$

This implies that $\frac{ds}{dw} - b_w < 0$.

As in my baseline model, mimickers will have a higher k_s,

$$\frac{dk_s}{dw} = k_{sE}E_w + k_{ss}s_E E_w > 0.$$

To analyse the behaviour of a mimicker differing in terms of a, I use a slightly different approach than above. Individuals now choose s and b while $E(s,b,K,a,r)$ is implicitly defined by the constraint $k(E,s,a) - rb = K$. The individuals problem and necessary conditions are

$$
\begin{aligned}
&\max_{\{s,b\}} U = u(B_1 - s + b) + \psi(B_2 + s - b(1+r)) + v(1 - Y/w - E(s,b,K,a,r)), \\
&U_s = -u' + \psi' - v'E_s = 0, \\
&U_b = u' - \psi'(1+r) - v'E_b = 0,
\end{aligned}
$$

where $E_s = -k_s/k_E$ and $E_b = r/k_E > 0$. The necessary conditions imply that $E_b + E_s < 0$. I will make use of the following derivatives

$$
\begin{aligned}
U_{ss} &= u'' + \psi'' - v'E_{ss} + v''E_s^2 < 0, \\
U_{bb} &= u'' + \psi''(1+r)^2 + v''E_b^2 < 0, \\
U_{sb} &= -u'' - \psi''(1+r) + v''E_sE_b > 0, \\
U_{sa} &= -v'E_{sa} + v''E_sE_a < 0, \\
U_{ba} &= -v'E_{ba} + v''E_bE_a < 0, \\
E_{ss} &= [k_{EE} + 2k_{Es}k_s - k_{ss}k_E]/k_E^2 > 0, \\
E_{bb} &= 0, \\
E_{sa} &= [k_{EE}E_ak_s - k_{sE}E_ak_E]/k_E^2 > 0, \\
E_{ba} &= -[rk_{EE}E_a + rk_{Ea}]/k_E^2 < 0.
\end{aligned}
$$

The effects of an increase in a on savings is negative

$$\begin{aligned}s_a\,|H| =& - U_{sa}U_{bb} + U_{ba}U_{sb}\\ =& v'u''(E_{sa}+E_{ba}) + v'\psi''(E_{ba}+(1+r)^2E_{sa})\\ &- v''u''E_a(E_s+E_b) - v''\psi''E_a(E_b+(1+r)^2E_s) < 0,\end{aligned}$$

where I made use of the following

$$\begin{aligned}(E_{sa}+E_{ba})k_E^2 =& k_{EE}E_a(k_s-r) + (k_{sE}k_aE_b - k_{Ea}k_sE_b) + k_{Ea}(k_sE_b+rE_s)\\ =& k_{EE}E_a(k_s-r) + k_{Ea}(k_sE_b+rE_s) > 0,\\ (E_{sa}E_b - E_{ba}E_s)k_E^2 =& k_{EE}E_a(k_sE_b+rE_s) + (k_{sE}k_aE_b - k_{Ea}k_sE_b) + k_{Ea}(k_sE_b+rE_s)\\ =& 0.\end{aligned}$$

The effects of an increase in a on b is ambiguous

$$\begin{aligned}b_a\,|H| =& - U_{ss}U_{ba} + U_{sb}U_{sa}\\ =& v'u''(E_{ba}+E_{sa}) + v'\psi''(E_{ba}+(1+r)E_{sa}) - (v')^2E_{ss}E_{ba}\\ &- v''u''E_a(E_b+E_s) - v''\psi''E_a(E_b+E_s).\end{aligned}$$

A mimicker will have a lower first period consumption,

$$(s_a - b_a)\,|H| = v'\psi''E_{sa}(1+r)r - v''\psi''E_aE_s((1+r)^2-1) + (v')^2E_{ss}E_{ba} < 0.$$

Thereby, mimickers will have a higher second period consumption. Finally, I need to establish that k_s is increasing in a

$$\begin{aligned}\frac{dk_s}{da} =& (k_{sa} - k_{sE}k_a/k_E) + k_{ss}s_a + k_{sE}(E_ss_a+E_bb_a)\\ =& k_{ss}s_a + k_{sE}(E_ss_a+E_bb_a) > 0,\\ (E_ss_a+E_bb_a)\,|H| =& v'u''(E_{ba}+E_{sa})(E_s+E_b) + v'\psi''[(E_{ba}+(1+r)E_{sa})E_b\\ &+ (E_{ba}+(1+r)^2E_{sa})E_s] - v''u''E_a(E_s+E_b)^2\\ &- v''\psi''[E_b(E_s+E_b) + E_s(E_b+(1+r)E_s)] - (v')^2E_{ss}E_{ba}E_b > 0.\end{aligned}$$

This shows that $MRS_c > M\hat{R}S_c$ and $k_s < \hat{k}_s$, therefore $(MRS_c - M\hat{R}S_c) + (1 - k_s/\hat{k}_s)(M\hat{R}S_c - 1) > 0$.