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http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/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 to Capital*

Arnaldur Sölvi Kristjánsson†

Abstract

This paper characterizes the optimal income and wealth tax schedules when rates of return are endogenous. Individuals exert investment effort in order to increase the return on their investments. Agents are heterogeneous along two dimensions: their investment ability and their 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, it is optimal to tax capital income and subsidize wealth. When wealth is not observed, it is optimal to tax capital income. The marginal tax rates on labour and capital income should not be equal, but they should be positively related to each other. The results extend to a model where individuals can hire investment advisors to increase the rate of return and also to a model with heterogeneous inheritance, in which case both capital income and wealth should be taxed.

JEL Codes: G11, H21, H24

Keywords: Optimal taxation, capital taxation, endogenous return to 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 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, with incomplete correlation between these two dimensions. The optimal tax system is therefore two-dimensional, there is 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 of the investment effort and ability. Therefore, capital income provides the government with information on 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.[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]

[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) for reviews and references of these various arguments.

The effect of heterogeneous returns, investment effort and investment ability has received almost no attention in the literature, beyond Stiglitz's (1985) and Gerritsen's et al. (2015) contributions.[3] Stiglitz (1985) sets up a very simple model where individuals differ in the rate of return they obtain, due to informational asymmetry. Gerritsen et al. (2015) set up a model where individuals are heterogeneous in their ability 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 of my approach is that the return to capital is homogenized, by making capital income depend on investment effort and investment ability as well as savings.

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. Second, differential rates of return may be due to uncertainty which is an inherent feature of financial markets.

There is an increasing empirical field analyzing the relationship between return and wealth. Two recent papers using Scandinavian administrative data find that returns are increase rapidly with wealth. Greengage et al. (2016) find that returns rise rapidly with wealth levels using Norwegian panel data. This is only partly driven by the wealthy taking more risk. Return differentials have a persistent component that explains almost 20% of the variation in returns. This is the main driver of the positive correlation between returns and wealth. Bach et al. (2015) perform a similar analysis using Swedish data. They also find a strong correlation between returns and wealth but this is primarily due to differences in risk taking. Differences in risk adjusted returns between households are significant but small.

[3] It should be noted that the concept that heterogeneity in investment abilities is relevant for optimal taxation is well acknowledged. For example, the textbooks of Kaplow (2008) and Salanié (2011) briefly mention the effects of heterogeneity in investment ability on optimal capital income taxation. Stiglitz notes that "[o]ne of the most important reasons for taxing capital income is that 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: 12).

There is also evidence from the US showing a positive correlation between returns and wealth. Piketty (2014) shows that returns 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 (2015) show that the same pattern emerges for the universe of U.S. foundations. Another piece of evidence from the Forbes global wealth rankings suggests that wealthier individuals tend to get higher returns (Piketty, 2014).

I set up a two period model of saving where individuals can exert investment effort, which increases the rate of return, and 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 ability and the possibility to make investment effort is a rational to tax capital income. Preferences are separable between consumption and leisure and thereby satisfy the condition for the Atkinson-Stiglitz theorem. I show that the zero tax on capital income is not optimal when one introduces the possibility to exert investment effort when individuals either differ in labour market productivity or investment abilities.

In my baseline model there are two different informational assumptions. In the first case the government observes wealth and capital income. In this case the government wants to downward distort the investment effort decision while keeping the savings decision undistorted. This is achieved by taxing capital income and subsidizing wealth. 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 unobserved but capital income is observed. In this case 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 save less and are therefore, at the margin, more willing to save. Also, more skilled individuals have, conditional on income, a higher rate of return. 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. First, I examine a model where individuals can hire financial advisors to increase their rate of return. The government observes wealth as well as capital income. In this extension, the government both wants to distort capital income as well as the intertemporal allocation. Interestingly though, I still get the result that capital income should be taxed and wealth subsidized. I also consider this model where individuals differ in their inheritance (instead of differing in their labour market productivity and investment ability). This is similar to the environment that Piketty (2014) argues to be appropriate. In this case, the government wants to tax wealth as well as capital income. In this model, the marginal tax rate on capital income exceeds the marginal tax rate on labour income.

The second extension is to add a domestic credit market allowing for interpersonal lending, assuming the government does not observe wealth. 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 tax 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 investing, which raises their return to 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 to 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. Individuals only invest in the project and do not participate in any way. 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 behavior of individuals will have no general equilibrium effects.

Capital income is denoted by $k(E, a, s)$ which is increasing in a and E and s, where s denotes savings and k_s is the return to savings. In addition, it has the following properties

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

The cross-derivative $k_{sa} > 0$ indicates that more able investors get a higher rate of return. $k_{Ea} > 0$ means that more able investors are more efficient, which is my definition of being an able investor. The cross-derivative k_{sE} is positive, meaning that the rate of return is increasing in investment effort. Since $k_{ss}, k_{EE} < 0$ there is decreasing return to savings and investment effort. The reason that $k_{ss} < 0$ is that investment projects are of finite size and with a given stock of knowledge (the combination of a and E) the rate of return will decline with s. Also, this ensures an interior solution.[4]

In the absence of the government, the budget constraint of individuals in period 1 and 2 are

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

where c_1 and c_2 are consumption 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 analyst 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 ob-

[4] Below, I solve optimal nonlinear taxes. This property is though not needed to get an interior solution if the government only has access to a linear capital income tax and a nonlinear labour income tax.

serve 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.[5] Estimating the market value of assets is on the other hand a much more daunting exercise. I believe that both informational assumptions are two extreme versions of reality and therefore analyst both cases.

The model assumes the typical asymmetric information between the government and individuals. 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 $a^h > a^l$.[6]

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

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

where n^i denotes the number of individuals of type i, B_1^i 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 denotes disposable income in the second period for

[5]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).

[6]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).

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 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.[7] The government's budget constraint, and also the resource constraint, in period 1 and 2 are, respectively

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

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

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

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. This ensures that the marginal utility of leisure equals the social gain of more labour supply as well as investment effort, i.e. $v_i' = \lambda_1 w^i = \lambda_2 k_E^i$. This means that more able individuals will be worse off.

By eliminating Lagrange multipliers, the optimal intertemporal and intratemporal

[7] 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.

allocations can be presented as

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

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

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

where MRS_Y^i denotes the marginal rate of substitution between labour income and present consumption. It shows how much an individual would need 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

$$\frac{k_E^i}{1 + k_s^i} = w^i,$$

which indicates that the marginal return to time use, in present value terms, is equated between L and E. The LHS shows the marginal benefit of increasing E measured in present value, and the RHS shows the marginal benefit of increasing L.

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

$$u(B_1^2-s^2)+\psi(B_2^2+s^2)+v(1-Y^2/w^2-E^2) \geq u(B_1^1-s^1)+\psi(B_2^1+s^1)+v(1-Y^1/w^2-\hat{E}^{21}), \tag{5}$$

where $\hat{E}^{21}$ is the investment effort chosen by a type 2 person mimicking a type 1 person, it is the value of E such that $K^1 = k(\hat{E}^{21}, s^1, a^2) = k(E^1, s^1, a^1)$. (5) ensures that a type 2 individual does not choose the bundle intended for a type 1 individual. A type 2 mimicker has the same consumption stream as a type 1 individual. Since a mimicker is more able than the type he is mimicking (either he has a higher a or w), he will have more leisure, i.e. $L^1+E^1 > \hat{L}^{21}+\hat{E}^{21}$. When individuals differ in terms of labour market productivity, then $Y^1/w^1 = L^1 > \hat{L}^{21} = Y^1/w^2$ and $E^1 = \hat{E}^{21}$. When individuals differ in terms of investment ability, then $Y^1/w = L^1 = \hat{L}^{21}$ and $E^1 > \hat{E}^{21}$.

4.1 Two type model: fixed investment ability

Here, I consider a two type model with fixed a where individuals differ in w, with $w^2 > w^1$. The government maximizes social welfare (2), subject to the government's revenue constraint (3) and the incentive constraint (5). In appendix A, the Lagrangian is presented, the necessary conditions derived and manipulated. The optimal allocation is

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

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

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

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

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

These conditions indicate that type 2 individual should be left undistorted. If implemented with a tax system, type 2 faces a marginal tax rates of zero. This is the standard no distortion at the top result as in the static Mirrleesian model and 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 is distorted downwards. According to (6c) the intertemporal allocation of type 1 should though be left undistorted. Also, 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. The tax system will therefore tax capital income but subsidize wealth. As is shown below, the marginal tax rate on labour income exceeds the rate on capital income. But the marginal tax rates are positively related to each other.

Individuals with higher w will, conditional on labour income, have more leisure because their labour supply is lower. Therefore, more productive individuals are, at the margin, also 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. But more productive individuals have more leisure, conditional on income, and therefore they are 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 independent of productivity, it only depends on labour and capital 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 willing to exert more investment effort and supply more labour. Therefore labour income and capital income should be used for taxation purposes. In addition to this, more able individuals have lower labour supply, conditional on income. 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. If the government is able to relax the incentive constraint it can perform more redistribution. Therefore, the government tries to find ways to relax the incentive constraint. 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. The government could either make one or more types better off without making any type worse off. The argument is analogous to the argument for distorting labour income downwards in the static Mirrlees model. The reason why the downward distortion relaxes the incentive constraint is that the type 2 mimicker has a lower marginal rate of substitution than type 1 individual, i.e. $MRS_K^1 > M\hat{R}S_K^{21}$, where $M\hat{R}S_K^{21} = \frac{\hat{v}'_{21}}{\hat{\psi}'_{21}\hat{k}_E^{21}}$ is the marginal rate of substitution for a mimicker.[8] 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 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 decision 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

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

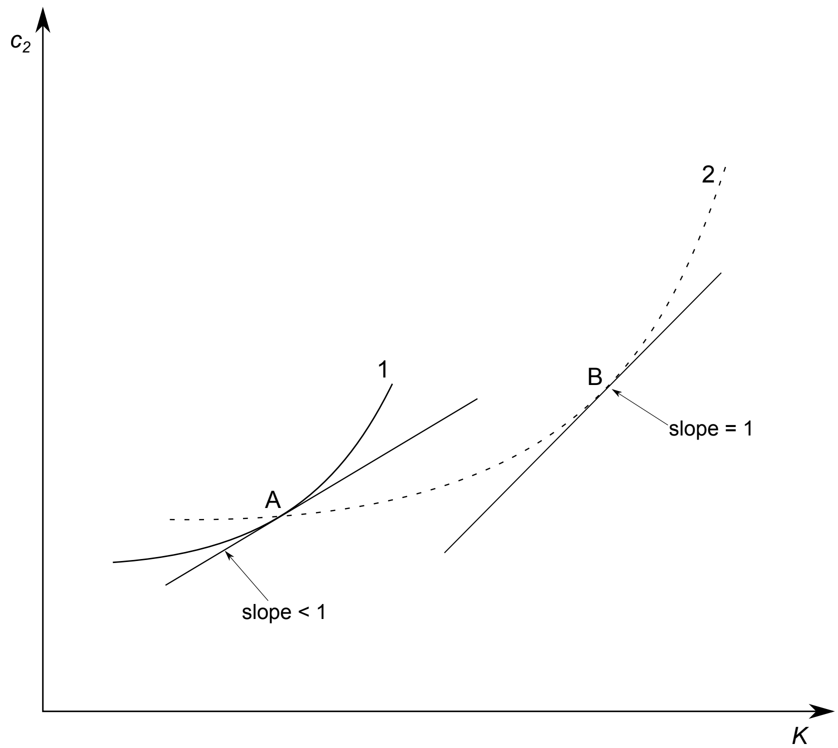

Figure 1: Indifference curves in K and c_2 space under optimal allocation.

income and a higher consumption in period 2, i.e. $K^2 > K^1$ and $c_2^2 > c_2^1$.

Condition (6d) shows that labour supply of the less productive worker should be distorted downwards, i.e. there should be a tax on labour income. This is just as in the static Mirrlees model. As in the discussion above, the downward distortion serves to relax the incentive constraint. A mimicker has more leisure than the type he mimics, because he has a higher labour market productivity and produces Y^1 with a lower investment effort than type 1, i.e. $L^1 > \hat{L}^{21} = L^1 \frac{w^1}{w^2}$. Since the mimicker has more leisure, he needs less compensation in terms of present consumption in order to supply one more unit of labour income, compared to the less able worker, i.e. $MRS_Y^1 > M\hat{R}S_Y^{21}$.

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), it follows that t', $T_K > 0$ and $T_s < 0$. This means that the marginal tax rate on capital income is positive while the marginal tax rate on savings will be negative. The marginal subsidy of savings ensures that the intertemporal allocation will be undistorted. In addition, it is shown that the marginal tax rate on labour income will be greater than for capital income, i.e. $t' > T_K$. I explain this intuitively below.

For the less productive type the marginal return is distorted between E and L as the marginal return (in present value) for L is higher than for E, i.e. $w^1(1 + k_s) > k_E^1$. The reason for this can be seen from the necessary conditions for L^1 and E^1, which are, respectively

$$n^1 v_1' - \gamma \hat{v}_{21}' \frac{w^1}{w^2} = n^1 \lambda_2 w^1 (1 + k_s^1),$$
$$n^1 v_1' - \gamma \hat{v}_{21}' = n^1 \lambda_2 k_E^1.$$

The LHS shows the cost of increasing L or E, while the RHS shows the benefit (which is the revenue effect of more L or E). The cost of increasing L or E is the utility loss for the less productive type net of utility loss for the mimicker. Mimickers are more efficient workers (as $w^2 > w^1$) but they are equally efficient investors (as $k_E(E^1, s^1, a) = k_E(\hat{E}^{21}, s^1, a)$). Therefore the utility loss for the mimicker of an increase in L is less than the utility loss of an increase in E. This means that the marginal cost of an increase in L is larger than for E. Since the government equates marginal costs and marginal benefits, the marginal benefit for L has to be greater than the marginal benefit of increasing E. In other words, $w^1(1+k_s^1)$ has to exceed k_E^1. This is achieved by having a lower marginal tax rate on capital income compared to labour income.

Since k_E is declining in E and the government wants to reduce the amount of E^1 (compared to an undistorted allocation), a decrease in the distortion on capital income below the distortion on labour income, creates an upward incentive for investment effort, which will reduce k_E^1. The relationship between MRS_Y and MRS_K can be written as

$$MRS_Y = 1 - t' = MRS_K^1 \frac{k_E^1}{w^1(1 + k_s)} < MRS_K^1 = 1 - T_K,$$

where $\psi_1'/\lambda_2 = u_1'/\lambda_1$. It follows from this relation that $t' > T_K$. 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 related to each other. This can be seen by writing the marginal tax rates for labour and capital income as, respectively

$$t' = \frac{\gamma u_1'}{n^1 \lambda_1}\left[MRS_K^1 - MR\hat{S}_K^{21}\frac{w^1}{w^2}\right]\frac{k_E^1}{w^1(1+k_s)},$$
$$T_K = \frac{\gamma u_1'}{n^1 \lambda_1}\left[MRS_K^1 - MR\hat{S}_K^{21}\right].$$

This shows clearly that a higher t' will be associated with a higher T_K.

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 the equations. 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 self selection 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. As in the discussion above, the downward distortion serves to relax the incentive constraint. A mimicker has more leisure than the type he mimics, because he is a more able investor and can produce K^1 with a lower investment effort than type 1, i.e. $E^1 > \hat{E}^{21}$.

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. At the optimum, the government wants to distort the marginal return between E and L for type 1, by having $w^1(1+k_s) < k_E^1$. This is achieved by setting $t' < T_K$. The results from this section are summarized in proposition 1.

Proposition 1 *If the government observes wealth, capital and labour income, the optimal tax system has a positive marginal tax rate on labour and capital income but* *a negative marginal tax rate on wealth. This will distort investment effort but keep the intertemporal allocation undistorted. The marginal tax rate for labour and capital income should not be equal, but positively related to each other. 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 observes 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), where K denotes capital income. 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). Properties of $s(E, a, K)$ are explored by substituting $s(E, a, K)$ into $k(\cdot)$ and then differentiating

$$k(E, s(E, a, K), a) - K = 0, \tag{8}$$

where $s_E = -k_E/k_s < 0$.

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

$$\max_{\{E\}} U = u(B_1 - s(E, a, K)) + \psi(B_2 + s(E, a, K)) + v(1 - Y/w - E). \tag{9}$$

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

$$U_E = -u' s_E + \psi' s_E - v' = 0. \tag{10}$$

Condition (10) 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 (10), it follows that the optimal choice of E is a function of all the exogenous variables, i.e. $E = E(Y, B_1, K, B_2, w, a)$. By implicitly differentiating (10), I can derive the derivatives of E w.r.t. w and a in order to analyst the behavior of the mimicker,

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

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

where $s_a = -k_a/k_s < 0$ and $s_{EE} = (k_{ss} s_E k_E + 2k_{sE} k_E - k_{EE} k_s)/k_s^2 > 0$. Note that in (11) and (12), Y, B_1, K, B_2 and either a or w are held constant. Therefore these derivatives indicate the behavior of the mimicker. The first ratio on the RHS in (11) is less than one and since $\frac{dL}{dw} = -\frac{Y}{w^2}$, it follows that $\frac{dE}{dw} < -\frac{dL}{dw}$. This means that a mimicker that has a higher w will choose a lower $E + L$ than the less productive type and will therefore have more leisure than the less able worker. It follows from (12) that mimickers who have higher a, will also have more leisure than the type being mimicked.

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 $s_{Ea} \geq 0$ then $\frac{dE}{da} > 0$, but if $s_{Ea} < 0$ the sign of $\frac{dE}{da}$ is ambiguous. If the following form of weak separability is satisfied

$$k(E, a, s) = F(f(E, s), a), \tag{13}$$

then $s_{Ea} > 0$. This will be assumed to be the case from now on.[9] 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.

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 $s(E(Y, B_1, K, B_2, w, a), a, K)$

$$\frac{ds}{dw} = s_E E_w < 0, \tag{14}$$

$$\frac{ds}{da} = s_E E_a + s_a < 0, \tag{15}$$

see appendix B for the sign of the inequality in (15), 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. Then an individual will find it beneficial to increase E. Due to the constraint $k(E, a, s) = K$, s will decrease. This follows from (11) and (14). 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 follows from (12) and (15). 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

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

derivatives of the indirect utility function follow from the envelope theorem

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

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

$$\begin{aligned} \max_{\{Y^i, B_1^i, K^i, B_2^i\}} \quad & \sum_i n^i V(Y^i, B_1^i, K^i, B_2^i, a^i, w^i) \\ \text{subject to} \quad & \sum_i n^i (Y^i - B_1^i) \geq g_1 \quad (\lambda_1), \\ & \sum_i n^i (K^i - B_2^i) \geq g_2 \quad (\lambda_2), \\ & V^2 \geq \hat{V}^{21} \quad (\gamma). \end{aligned} \tag{16}$$

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 (16).

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

$$MRS_K^1 = 1 - \frac{\gamma \hat{\psi}'_{21}}{n^1 \lambda_2} \left[MRS_K^1 - M\hat{R}S_K^{21} \right] < 1, \tag{17a}$$

$$MRS_c^1 =(1+k_s^1) - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{21}) + (1-k_s^1/\hat{k}_s^{31})(M\hat{R}S_c^{21}-1)\right] \tag{17b}$$
$$< 1+k_s^1,$$
$$MRS_Y^1 =1 - \frac{\gamma\hat{u}'_{21}}{n^1\lambda_1}\left[MRS_Y^1 - M\hat{R}S_Y^{21}\right] \lesseqgtr 1. \tag{17c}$$

Condition (17b) 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. The reason is that this serves to relax the binding incentive constraint and thereby the government can achieve a Pareto improvement. The government could either make one or both types better off without making any type worse off. 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.

An important feature of the distortion on capital income from condition (17b) 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. The only thing that matters for the distortion is that mimickers have more leisure than the less productive worker. Whether the worker has zero or positive E is not relevant for the result that capital income should be taxed.

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 back-load consumption. This means that the intertemporal allocation provides the government with information on labour market productivity and should be used for taxation purposes.

There are two reasons why the downward distortion relaxes the binding 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 (17b). 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 (17b).

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.

$MRS_c^1 > \hat{MRS}_c^{21}$. 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 (since he has less (more) present (future) 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 3). 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.

Suppose one starts from an undistorted allocation satisfying the incentive constraint, where $MRS_c^1 = \lambda_1/\lambda_2$. Consider a small variation along the indifference curve with $dB_2^1 < 0$ along with a variation $-dB_1^1 \cdot MRS_c^1 = dB_2^1 < 0$, which has no effect on the utility of type 1. But type 2 mimicking type 1 is not at an undistorted allocation before the small variation since $MRS_c^1 > \hat{MRS}_c^{21}$. The small variation will therefore decrease utility of type 2 when mimicking (type 2 would be indifferent only if $dB_2^1 < -dB_1^1 \frac{\lambda_1}{\lambda_2}$). This small variation will therefore relax the incentive constraint.

A different way of viewing the existence of the distortion is the following. As in the model where the government also observed s, investment effort should be distorted downwards, see (17a). Since the government does only observe K, it is impossible to have $MRS_K^1 < 1$ and $MRS_c^1 = 1 + k_s^1$ at the same time. Therefore the government is compelled to set $MRS_c^1 < 1 + k_s^1$.

The optimal allocation for type 1 and 2 are shown graphically in figure 2, which resembles closely the static Mirrleesian model.[10] 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.

Condition (17c) 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

[10]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.

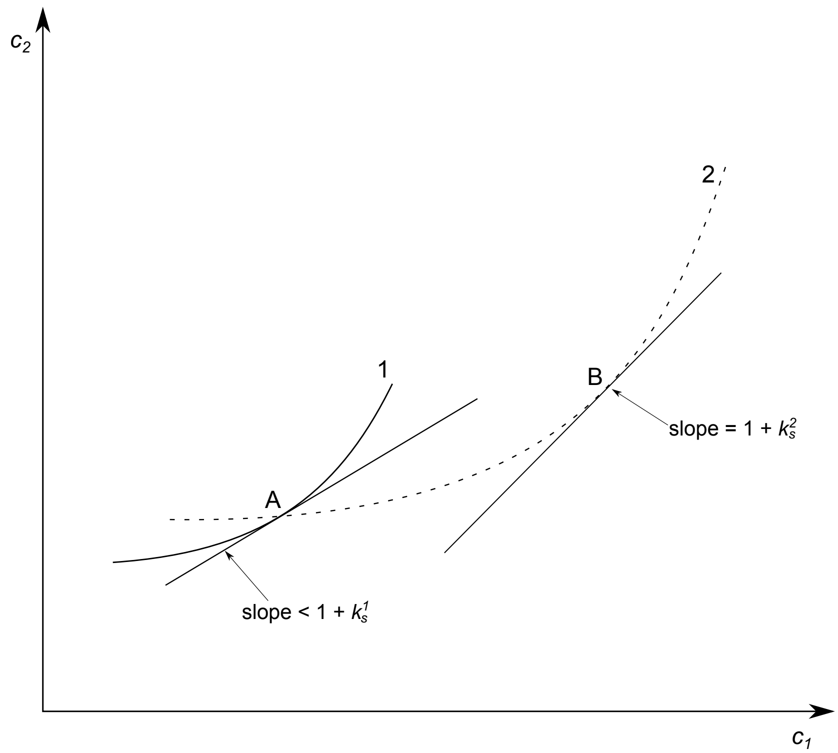

Figure 2: Indifference curves in c_1 and c_2 space under optimal allocation.

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 (17c) is ambiguous.

I want to know whether the distortion on labour income exceeds, falls short of or is equal to the distortion on capital income. From the necessary conditions shown in appendix B, it can be seen that the government wants to distort the marginal return for L and E in favor of L. I want to know whether this translates into marginal tax rates being higher for capital income than labour income (as was the case in section 4.1). Since individuals receive labour income in 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 received in the second period. Individuals have an homogeneous endowment e in period 1 and face the following budget constraint

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

In appendix B, I show that the marginal tax rate on labour income exceeds the marginal tax rate on capital income. The reason is that at the optimum the government wants to distort the optimal mix between E and L such that $w^1 > k_E^1$. This will be achieved through a tax system that sets a higher marginal tax rate on labour income than on capital income. The intuition behind this results is analogous to the discussion in the end of section 4.1. In order to reduce k_E^1 below w^1, the government needs to increase the incentive for investment effort since k_E^1 is declining in E. This is done by setting a lower distortion on capital income. 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.

5.3 Two type model: fixed labour market productivity

I analyst the same problem as in (16), 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 (17a-17c). The rationale and intuition for the distortions are identical, they serve to relax the incentive constraint.

As in the model with heterogeneous w, there is a downward distortion on capital income. The rationale is completely 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 a more effective investor (see appendix B). Due to these two differences a downward distortion will relax the incentive constraint.

Turning to the labour-leisure allocation for type 1 individuals in (17c), it is not possible to determine the sign of optimal distortion (or if by pure chance there should be no distortion). The reason is that there are two counteracting forces. First, because less able investors save more and exert less investment effort than the mimickers, they need, all else equal, a smaller compensation in terms of present consumption in order to produce one more unit of Y. This implies an upward distortion. Second, mimickers have more leisure (as they need less investment effort to mimic type 1) and therefore they need less compensation in terms of present consumption to produce one more unit of Y. This calls for a downward distortion of labour supply.

Turning to the difference in distortions between labour and capital income. From the governments necessary conditions (shown in appendix B) it can be seen that the government wants to distort the marginal return for L and E in favor of E. As in section 5.2, it is uncertain whether this means that the marginal tax rate will be higher or lower 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 and the intuition exactly opposite and no need to repeat the discussion here. 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. The marginal tax rate for labour and capital income should not be equal, but related to each other. 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 analyst the model with four types of individuals. The distribution of types is shown in figure 3. Analyzing the complete four type model adds complications to the problem. But it is nonetheless possible to get some analytical results without relying on numerical examples. 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. But the direction of redistribution between type 3 and 2 depends on the joint distribution of w and a. 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). 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.

Even if I know the direction of the incentive constraint, it is not possible to characterize the behavior of the mimicker of type 3 or 2 in general. Therefore it is not possible to sign the direction of the distortion. To sign the distortion, I need to know the direction of the following two inequalities: $s^i \gtreqless \hat{s}^{ji}$ and $k^i_s \gtreqless \hat{k}^{ji}_s$. It cannot be

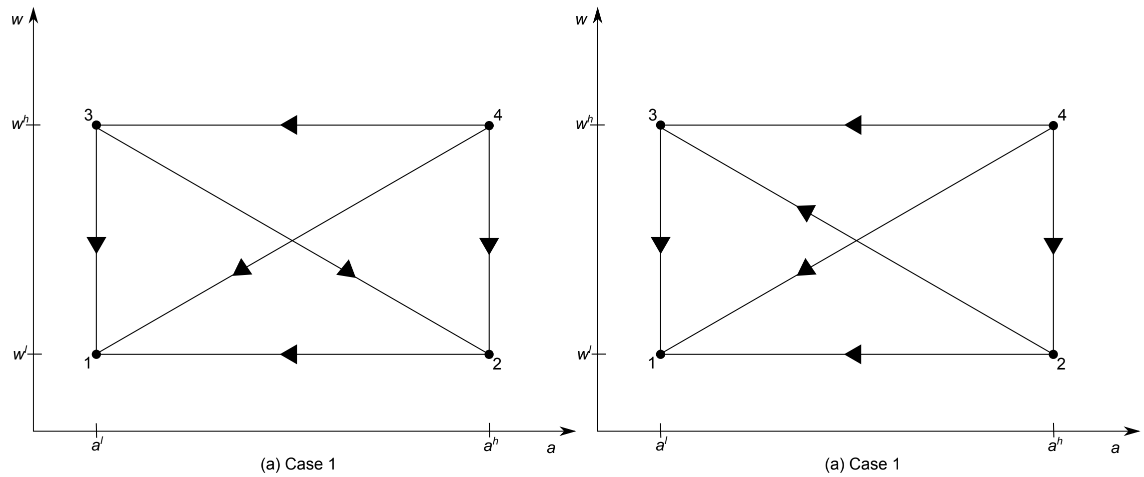

Figure 3: Potentially binding incentive constraints in case 1 and 2.

shown in general whether type 3 (2) mimicking type 2 (3) will save more, i.e. $\hat{s}^{32} \lesseqgtr s^2$ ($\hat{s}^{23} \lesseqgtr s^3$), and have a higher rate of return or not, i.e. $\hat{k}_s^{32} \lesseqgtr k_s^2$ ($\hat{k}_s^{23} \lesseqgtr k_s^3$). The reason is that 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 and type 2 the opposite.

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

$$\begin{aligned}\mathcal{L} = & \sum_i^4 n^i V^i + \gamma^{21}\left[V^2 - \hat{V}^{21}\right] + \gamma^{31}\left[V^3 - \hat{V}^{31}\right] \\ & + \left(\gamma^{32}\left[V^3 - \hat{V}^{32}\right] + \gamma^{23}\left[V^2 - \hat{V}^{23}\right]\right)\gamma^{42}\left[V^4 - \hat{V}^{42}\right] + \gamma^{43}\left[V^4 - \hat{V}^{43}\right] \\ & + \lambda_1\left[\sum_i^4 n^i(Y^i - B_1^i) - g_1\right] + \lambda_2\left[\sum_i^4 n^i(K^i - B_2^i) - g_2\right],\end{aligned} \tag{18}$$

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 analyzing 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). The necessary conditions are derived in appendix C and the optimal intertemporal allocations are shown there as well. From the four type model two general results are obtained, which are summarized in proposition 3. The underlying mechanism is the same as in the two type model.

Proposition 3 *If the government observes labour and capital income, but not wealth, the optimal tax system has the following structure, depending on the two cases depicted in figure 3. Of type 2 and 3 one should face a downward distortion and one face an ambiguous distortion. In case 1(2), type 2(3) should have a positive marginal tax rate on capital income. Type 1 should always have a positive marginal tax rate on capital income.*

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

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

where $k(\cdot)$ has the following properties[11]

$$k_{Em} \geq 0,\ k_m, k_{sm} > 0,\ k_{mm} < 0.$$

I assume that E and m are not substitutes, they can be either complements ($k_{Em} > 0$) or unrelated phenomenons ($k_{Em} = 0$). In the case of complementarity between E and m one could think that financial advisors give individuals some information about investment options while individuals need to make own choices to some extent. On the contrary substitutability could be argued if it is the total effort of individuals and financial advisors that matter, but I rule out this case. Buying financial information leads to an increase in capital income, but at a decreasing rate. Also, an increase in m leads to a higher rate of return, hence $k_{sm} > 0$.

I will consider optimal allocations in a two type model, where individuals will either differ in w or a. The setup is essentially the same as in section 4. The government observes Y, s and K, where individuals can deduct expenses on financial advisors from the capital income tax base, i.e. the government observes $K = k(m, \cdot) - m$.[12] 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 (either having lower w or a).

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 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 means that the simple fact

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

[12] The results would be quantitatively unchanged if m where not deductible, i.e. if the government would observe $K = k(E, m, s, a)$. The important assumption is that the government does not observe m.

that individuals have heterogeneous returns, which might be positively correlated with wealth, 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 taxed in this model.

In case 2, a different result emerges. In the case when 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. If individuals differ in terms of investment ability, wealth should be distorted downwards as well as capital income. The optimal intertemporal allocation is characterized by the following condition

$$\begin{aligned} MRS_c^1 =& 1 + k_s^1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[MRS_c^1 - M\hat{R}S_c^{21}\right] < 1 + k_s^1 \\ =& k_m^1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[(MRS_c^1 - 1) - (M\hat{R}S_c^{21} - 1)\frac{k_m^1 - 1}{\hat{k}_m^{21} - 1}\right] < k_m^1. \end{aligned} \tag{19}$$

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. If I consider consumption in the second period in terms of a tax function,

$$c_2 = s + k(m, s, a) - T(s, k(m, s, a)).$$

It follows from (19) that $T_K > 0$ and $T_K k_s + T_s > 0$, while T_s has an ambiguous sign.[13] Even though it can be shown that the intertemporal allocation should be distorted downwards, the sign of the marginal tax rate on wealth is ambiguous.

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 has the following optimality conditions

$$MRS_K^1 = 1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[MRS_K^1 - M\hat{R}S_K^{21}\right] < 1, \tag{20a}$$

$$MRS_c^1 = 1 + k_s^1 - \frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\left[MRS_c^1 - M\hat{R}S_c^{21}\right] < 1 + k_s^1. \tag{20b}$$

These conditions are for both a model where individuals differ in w and a (with the same direction of inequality in both models). These conditions show that both capital income as well as savings should be distorted downwards. From condition (20a) it follows that

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

the marginal tax rate on capital income will be positive. Since capital income is taxed, it is not immediately clear from (20b) that wealth should also be taxed. In appendix D, it is shown that wealth should be subsidized, i.e. $T_s < 0$. The results are summarized in proposition 4.

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

1. $k(m, s)$*: When there is heterogeneity in* w*, there is no taxation of wealth nor capital income.*

2. $k(m, s, a)$*: When there is heterogeneity in* w*, wealth and capital income should not be taxed. When there is heterogeneity in* a*, savings should be distorted downwards. In this case the marginal tax rate on capital income is positive but the marginal tax rate on wealth is ambiguous.*

3. $k(E, m, s, a)$*: When there is heterogeneity in* w *or* a*, investment effort and savings should be distorted downwards. The marginal tax rate on capital income is positive but the marginal tax rate on wealth is negative.*

6.2 Buying investment information with initial endowment

Piketty (2014) argues that a large part of inequality is based on inheritance. Also he argues that rates of return are increasing in initial endowments since individuals with large wealth are able to spend more money on financial advisors. Therefore he predicts wealth inequality to increase in future, particularly at the top. For this reason he argues for an annual progressive wealth tax. The results I have found so far do not give a rational for subsidizing taxing wealth, the results have found arguments for subsidizing wealth. Here I analyst how the results would alter if individuals have heterogeneous initial endowment.

In this model individuals start with an initial endowment, denoted by e, which is interpreted as inheritance received. A more realistic modeling strategy is to endogenously determine inheritance. Then, the bequeather as well as the inheritor would benefit from the inheritance and the benefits of inheritance would be counted twice. Taking this into account generally weakens the case for taxing capital income (see e.g. Boadway et al., 2000 and Cremer et al., 2003). For simplicity, the benefit of the bequeather will not be taken into account.

As above, individuals can increase their rate of return by spending money on financial advisors. To emphasize on the core idea, capital income does not depend on investment ability. Capital income is now $k(E, s, m)$ with assumption as discussed above.

In the absence of the government the individual's budget constraint is

$$c_1 = Y - s + e - m,$$
$$c_2 = s + k(E, s, m).$$

I set up two type model where individuals only differ in their initial endowment, with $e^2 > e^1$. The government observes Y, s and K but not e. As above, expenditure on m is tax deductible, i.e. $K = k(E, s, m) - m$. In reality governments do observe inheritance (even though avoidance problems might be more severe than with other tax bases). If the government could observe e, they could simply tax it away without creating any distortions. With endogenously determined inheritance, this would not be the case.

I further explain the model in appendix E, where the government's problem is solved. The optimal allocation resulting from this model take the same form as in (20a) and (20b), with the same inequalities. This means that the government wants to downward distort the intertemporal allocation as well as the investment effort, i.e. capital income as well as wealth should be distorted downwards. In terms of the tax system, there is a somewhat different optimal system than in case 3 in section 6.2. In appendix E, I show that the optimal tax system will be such that the marginal tax rate on capital income as positive ($T_K > 0$) as well as the marginal tax rate on wealth ($T_s > 0$). This means that wealth as well as capital income should be taxed. This should not come as a surprise as the government wishes to redistribute from the rich to the poor. While in the baseline model the government redistributes from those with high skill to those with low skill.

I want to know whether the marginal tax rate on labour income is larger or smaller than the marginal tax rate on capital income. To do so I have to follow a similar procedure as in section 5.2, and consider a model where labour income and capital income are both received in the second period. Otherwise, the setup is unchanged. In such a model, the marginal tax rate on capital income exceeds the marginal tax rate on labour income. This is shown in appendix E. The intuition is similar to the one in section 4.1. From the government's point of view, the marginal cost of increasing E^1 is higher than the marginal cost of increasing L^1. This is because mimickers respond with a smaller change in investment effort than the poor individual (type 1) in response to an increase in E^1 (because $k_E^1 < \hat{k}_E^{21}$). Meanwhile, mimickers respond with exactly the same change in labour supply than the poor individual in response to an increase in L^1. Therefore the benefit of increasing E^1 (which is $\lambda_2 n^1 k_E^1$) has to exceed the marginal benefit of increasing L^1 (which is $\lambda_2 n^1 w$). This implies that $k_E^1 > w$, which means that the tax system distorts the marginal return to E and L. A tax system that equates the marginal time use between E and L sets equal marginal tax rates on labour income and capital income. Since k_E is declining in E, a higher marginal tax rate on capital income than on labour income will reduce investment effort and lead to $k_E^1 > w$. The results are summarized in proposition 5.

Proposition 5 *When individuals can spend money on financial advisors and individuals differ in terms of their initial endowment and the government observes wealth, capital and labour income the optimal allocation is such that investment effort and the savings decision is distorted downwards. In this case the marginal tax rate on capital income and wealth are both positive. When individuals receive labour and capital income in the same period, the marginal tax rate on capital income will exceed the rate on labour income.*

6.3 Domestic credit market

In the models I have considered so far, the domestic economy faces a market imperfection since there was no existing 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). In the absence of the government, the budget constraint of an individual is

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

where b is the amount borrowed (if $b > 0$) or lend (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 there is perfect loss offset of interest payments. 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, government's do have some knowledge on the nature of capital income. But if the government wants to tax different forms of capital income homogeneously to preserved neutrality, the assumed informational structure might be 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

$$\sum n^i b(Y^i, B_1^i, K^i, B_2^i, r, w^i, a^i) = 0. \tag{21}$$

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 homogenous, i.e. $r = k_s$ (this may not be the case with taxation).

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).[14] When K increases, individuals will mechanically save more in order for $k(E, s, a) - rb$ to equal

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

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 (note that an increase in s (b) will decrease (increase) c_1).

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 (21). 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 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 $\frac{\partial b}{\partial B_2} > 0$. All the comparative statics are shown in appendix F. 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, 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

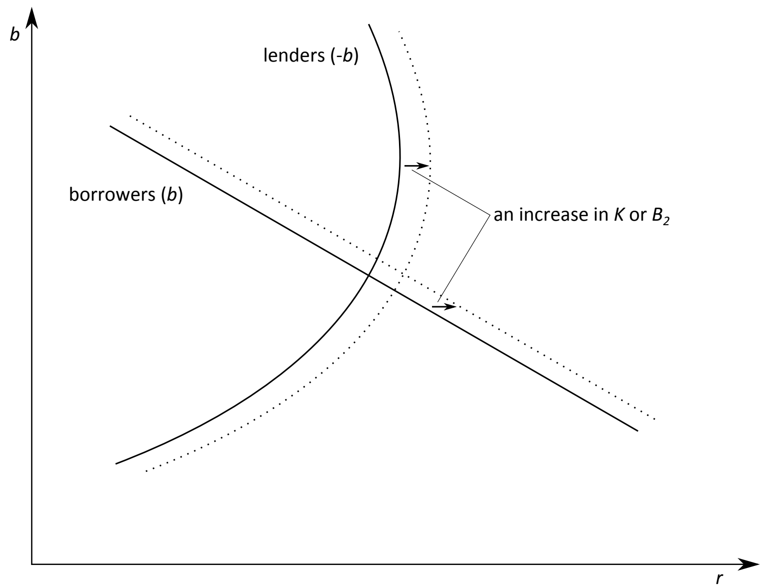

Figure 4: Effects of a raise in K or B_2 on r.

or lenders. For borrowers, the income effects are negative as a raise 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 4 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 lending 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 4.

Now I can analyst 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 raise in K or B_2 will shift the demand curve downwards. Since they now want to lend less at a given r.

From figure 4, it can be seen that if K or B_2 increases for either the borrower or lender, the equilibrium r will increase.[15] This means that $\frac{\partial r}{\partial K^i} > 0$ and $\frac{\partial r}{\partial B_2^i} > 0$ for all

[15]The formal analysis behind the comparative statics in figure 4 is done by totally differentiating condition (21) w.r.t. K and B_2:

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

types. This is an important fact for the results below.

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 F 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 F, I derive the necessary conditions for the government's problem. The optimal intertemporal allocation is

$$\begin{aligned} MRS_c^1 =&(1+k_s^1) - \frac{\gamma\hat{\psi}_{21}}{\lambda_2 n^1}\left[(MRS_c^1 - M\hat{R}S_c^{21}) + \left(1 - \frac{k_s^1}{\hat{k}_s^{21}}\right)(M\hat{R}S_c^{21} - 1)\right] \quad (22a)\\ &- \frac{\theta}{\lambda_2 n^1}\left[\frac{\partial r}{\partial K^1}k_s^1 + \frac{\partial r}{\partial B_2^1}(MRS_c^1 - 1)\right],\\ MRS_c^2 =&(1+k_s^2) - \frac{\theta}{\lambda_2 n^2}\left[\frac{\partial r}{\partial K^2}k_s^2 + \frac{\partial r}{\partial B_2^2}(MRS_c^2 - 1)\right], \quad (22b) \end{aligned}$$

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 (22a) are positive as well as the bracket in (22b) 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.

The second term on the RHS in condition (22a) is an identical term as appears in condition (17b). In appendix F, it is shown that the term is positive leading to a downward distortion, just as in (17b). 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 raise in r is 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

---

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

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 4.

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 raise 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 raise 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 (22a) and the second term in (22b) 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. 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. The results are summarized in proposition 6.

Proposition 6 *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 term implies a positive marginal tax rate if a rise in r is socially undesirable, 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, depending on the 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 at the optimum. 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 higher labour market productivity have, conditional on income, lower labour supply and are at the margin more willing to work. This means that labour income will provide the government with information about individual productivity. In my model, the same holds for capital income, more skilled individuals are, conditional on income, at the margin more willing to save and exert investment effort. 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 is the marginal tax rate on labour income, the higher should the marginal tax rate on capital income be.[16] Which marginal tax rate should be higher is in general ambiguous. The results indicate that as the heterogeneity of labour market productivity is larger, compared to the heterogeneity of investment ability or initial endowments, the marginal tax rate on labour income should be larger.

A distinct feature of the model is that a capital income tax and a wealth tax have different functions. I know of no other optimal tax model that has such a feature.[17] In a model with perfect capital markets it makes no difference whether the government taxes capital income or wealth. For example, a 30% tax on capital income with a return of 5% is equivalent to a 1,5% tax on wealth. This is not the case in my model. In all the models with heterogeneous labour market productivity and heterogeneous investment ability, capital income should be taxed while wealth should be subsidized. In section 6.2, I consider a model where individuals only have heterogeneous endowments and individuals can hire financial advisors. In that model, the government would want to tax both capital income and wealth. It should come as no surprise as a surprise as the government wishes to redistribute from the rich to the poor. This is in contrast to the baseline model the government redistributes from those with high skill to those with low skill. This means that the optimal way to design the taxation of capital crucially

[16]This is in line 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."

[17]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 behavioral effects if tax policy is time consistent.

depends on various features of individuals, the capital market and the information available to the government.

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 intrapersonal 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 a 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.

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.

Boadway, R., Marchand, M. and Pestieau, P. (2000). Redistribution with Unobservable Bequests: A Case for Taxing Capital Income. *The Scandinavian Journal of Economics*, 102(2), 253-267.

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

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.

Cremer, H., Pestieau, P. and Rochet, J.-C. (2003). Capital income taxation when inherited wealth is not observable. *Journal of Public Economics*, 87(11), 2475-90.

Greengage, A, Guiso, L., Malacrino, D. and L. Pistaferri (2016). "Wealth returns dynamics and heterogeneity", Stanford University mimeo.

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

Kaplow, L. (2008). *The Theory of Taxation and Public Economics*. Princeton and Oxford: Princeton University Press.

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.

Salanié, B, (2011). *The Economics of Taxation.* Cambridge, MA: MIT Press.

Saez, E. and Zucman, G. (2015). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *Quarterly Journal of Economics*, forthcoming.

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

Slemrod, J. and Gillitzer, C. (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 (ed.), *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.

Appendix

A Calculations for section 4

Necessary conditions In section 4 there are two different models. Both models have the same form and they would 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

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

where $U^i = u(B_1^i - s^i) + \psi(B_2^i + s^i) + v(1 - Y^i/w^i - E^i)$ and $\hat{U}^{21} = u(B_1^1 - s^1) + \psi(B_2^1 + s^1) + v(1 - Y^1/w^2 - \hat{E}^{21})$. The necessary conditions are

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

where I used $\frac{d\hat{E}^{21}}{dE^1} = \frac{k_E^1}{\hat{k}_E^{21}}$, which follows from $k(E^1, a^1, s^1) - k(\hat{E}^{21}, a^2, s^1) = 0$. The optimal allocation in (6b), (6c) and (6e) follow immediately. To get (6a), I solve for $n^1 v_1'/k_E^1$ and $n^1 \psi_1'$ in the necessary conditions for E^1 and B_2^1, respectively, and divide them together,

$$\frac{v_1'}{\psi_1' k_E^1} = \frac{\gamma \hat{v}_{21}'/\hat{k}_E^{21} + \lambda_2 n^1}{\gamma \psi_1' + \lambda_2 n^1},$$

multiplying 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, $k_E^1(E^1, a^1, s^1) < \hat{k}_E^{21}(\hat{E}^{21}, a^2, s^1)$, as $E^1 > \hat{E}^{21}$. These two facts imply that $MRS_K^1 = \frac{v_1'}{\psi_1' k_E^1} > M\hat{R}S_K^{21} = \frac{\hat{v}_{21}'}{\hat{\psi}_{21}' \hat{k}_E^{21}}$. It also follows that $MRS_Y^1 = \frac{v_1'}{u_1' w^1} > M\hat{R}S_Y^{21} = \frac{\hat{v}_{21}'}{\hat{u}_{21}' w^2}$ in both models since $w^2 \geq w^1$.

B Calculations for section 5

Deriving (15) To derive the inequality in (15), I differentiate (10) to determine E_a and use (8) to determine s_E and s_a. By some manipulation the inequality follows

$$
\begin{aligned}
\frac{ds}{da}U_{EE} =& - U_{Ea}s_E + s_a U_{EE} \\
=& s_E(u' - \psi')s_{Ea} - s_E^2 s_a(\psi'' + u'') - s_a(u' - \psi')s_{EE} + s_a s_E^2(\psi'' + u'') + s_a v'' \\
=& \frac{u' - \psi'}{k_s^2} s_a \left[k_{EE}k_s - k_{sE}k_E\right] + s_a v'' > 0 \implies \frac{ds}{da} < 0.
\end{aligned}
$$

Necessary conditions The Lagrangian for (16) 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

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

Intertemporal allocation Solving the necessary conditions for K^1 and B_2^1 for $n^1 u_1$ and $n^1\psi_1$, and dividing, I obtain

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

Multiplying both sides with $(\gamma\hat{\psi}'_{21}+\lambda_2 n^1)/\lambda_2 n^1$ and rearranging yields

$$\frac{u'_1}{\psi'_1}=\frac{\psi'_1}{\lambda_2}+k_s^1-\frac{u'_1}{\psi'_1}\frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}+\frac{\gamma\hat{\psi}'_{21}}{n^1\lambda_2}\frac{k_s^1}{\hat{k}_s^{21}}\left(\frac{\hat{u}'_{21}}{\hat{\psi}'_{21}}-1\right).$$

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 (17b).

In order to show the inequality in (17b), 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. Here, $E(\cdot)$ is the individual's optimal choice, which is analyses by differentiating (10) and $s(\cdot)$ (which follows from the constraint (8)),

$$\frac{dk_s}{da}=k_{sE}E_a+k_{ss}(s_E E_a+s_a)+k_{sa},$$

where

$$\begin{aligned}
E_a=&\frac{-U_{Ea}}{U_{EE}}=\frac{(u'-\psi')s_{Ea}-(u''+\psi'')s_E s_a}{-(u'-\psi')s_{EE}+(u''+\psi'')s_E^2+v''}<0,\\
s_{Ea}=&\frac{k_E}{k_s}\left(\frac{k_{sa}}{k_s}-\frac{k_{Ea}}{k_E}\right)+\frac{s_a k_{ss}k_E-k_{Es}s_a k_s}{k_s^2}=\frac{s_a k_{ss}k_E-k_{Es}s_a k_s}{k_s^2}>0,\\
s_{EE}=&\frac{k_{ss}s_E k_E+k_{sE}k_E-k_{sE}s_E k_s-k_{EE}k_s}{k_s^2}>0,
\end{aligned}$$

where under the from (13) $\frac{k_{sa}}{k_s}=\frac{k_{Ea}}{k_E}$. Plugging into $\frac{dk_s}{da}$, I get after some manipulation

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

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. If a stronger separability in $k(E,s,a)$ is assumed than has been already done,

$$k(E,s,a) = f^1(E)f^2(s)g(a)$$

then $k_{sE}k_a - k_{sa}k_E = 0$ and then $\frac{dk_s}{da}U_{EE} < 0$, which implies that $\frac{dk_s}{da} > 0$. I will assume this to hold. 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 (10), 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 (17b) is positive. This proves the inequality in (17b).

Showing that the inequality in (17b) 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 (10) and $s_E < 0$ from (8).

Intratemporal allocation To solve the optimal intratemporal allocation, I solve the necessary conditions for $Y^1$and B_1^1 for $n^1 v_1$ and $n^1 u_1$ and divide

$$\frac{v'_1}{u'_1} = \frac{\gamma\hat{v}'_{21}\frac{w^1}{w^2} + \lambda_1 n^1 w^1}{\gamma\hat{u}'_{21} + \lambda_1 n^1}.$$

Multiplying both sides and rearranging yields

$$\frac{v'_1}{u'_1}\frac{n^1\lambda_1}{\gamma\hat{u}'_{21}} = \frac{\gamma\hat{u}'_{21}}{n^1\lambda_1}w^1 - \left[\frac{v'_1}{u'_1} - \frac{\hat{v}'_{21}}{\hat{u}'_{21}}\frac{w^1}{w^2}\right].$$

Multiplying both sides will give condition (17c).

In order to derive the inequality in (17c) for the model with heterogeneous a, I note that $s^1 > \hat{s}^{21}$, $E^1 > \hat{E}^{21}$ and $L^1 = \hat{L}^{21}$. This means that $u'_1 > \hat{u}'_{21}$ and $v'_1 < \hat{v}'_{21}$. Therefore, the direction of the inequality in (17c) cannot be signed.

In the model with heterogeneous w, I note that $s^1 > \hat{s}^{21}$, $E^1 < \hat{E}^{21}$, $L^1 > \hat{L}^{21}$ and $L^1 + E^1 > \hat{L}^{21} + \hat{E}^{21}$. This means that $u'_1 > \hat{u}'_{21}$, but I cannot sign $v'_1 > \hat{v}'_{21}$. Since $\frac{1}{u'_1} < \frac{1}{\hat{u}'_{21}}$, $v'_1 > \hat{v}'_{21}$ and $\frac{w^1}{w^2} < 1$ the direction of the inequality in (17c) is ambiguous for the model with heterogeneous w.

Difference in distortion between Y and K I consider the difference in distortion between Y and K starting with the model where individuals differ in their labour market productivity. 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 (9). All the comparative statics are analogous in this model and no need to repeat them here. The government's budget constraint is now

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

In the first best, I have the same optimal allocations as in (4b) and (4c), while condition (4b) is now

$$\frac{v_i'}{\psi_i'}\frac{1}{w^i} = MRS_Y^i = 1.$$

When setting up the government's problem, I follow a slightly different approach than above. Individuals choose s freely, while E adjusts so that the constraint $k(E, s, a) = K$ holds. This will give, due to the envelope theorem, $\frac{\partial V}{\partial K} = -\frac{v'}{k_E}$. The government's necessary conditions for K^1 and Y^1 are (the necessary conditions for B_1^i and B_2^i are the same as above)

$$(K^1) \qquad -n^1 v_1'/k_E^1 + \gamma \hat{v}_{21}'/\hat{k}_E^{21} + \lambda_2 n^1 = 0,$$
$$(Y^1) \qquad -n^1 v_1'/w^1 + \gamma \hat{v}_{21}'/w^2 + \lambda_2 n^1 = 0.$$

Combining these conditions, I get $w^1 > k_E^1$, the result follows from the fact that $k_E^1 > \hat{k}_E^{21}$, since $\frac{dk_E}{dw} = k_{EE}E_w + k_{Es}s_E E_w < 0$. In other words, the government wants to distort the marginal return for L and E in favor of L. The governments optimal allocation in this model is (performing manipulations identical to the ones above)

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

Next, I rewrite MRS^1_Y to show that $T_Y > T_K$

$$MRS_Y = 1 - t' = MRS^1_K \frac{k^1_E}{w^1(1+k_s)} < MRS^1_K = 1 - T_K \;\longrightarrow\; T_Y > T_K.$$

For the model with heterogeneous a, I have exactly the opposite result. This is since $k^1_E < \hat{k}^{21}_E$ since

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

since $U_{EE} < 0$, then $\frac{dk_E}{da} > 0$. Here, $E(\cdot)$ follows from (10) and $s(\cdot)$ from (8).

C Calculations for section 5.3

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 (18) are

$$\begin{aligned}
&(Y^1) && - n^1 v'_1/w^l + \gamma^{21}\hat{v}'_{21}/w^l + \gamma^{31}\hat{v}'_{31}/w^h + \lambda_1 n^1 = 0, \\
&(Y^2) && - (n^2 + \gamma^{21} + \gamma^{23})v_2/w^l - \gamma^{32}\hat{v}'_{32}/w^h - \gamma^{42}\hat{v}'_{42}/w^h + \lambda_1 n^2 = 0, \\
&(Y^3) && - (n^3 + \gamma^{31} + \gamma^{32})v'_3/w^h - \gamma^{23}\hat{v}'_{23}/w^l - \gamma^{43}\hat{v}'_{43}/w^h + \lambda_1 n^3 = 0, \\
&(K^1) && - n^1(u'_1 - \psi'_1)/k^1_s + \gamma^{21}(\hat{u}'_{21} - \hat{\psi}'_{21})/\hat{k}^{21}_s + \gamma^{31}(\hat{u}'_{31} - \hat{\psi}'_{31})/\hat{k}^{31}_s + \lambda_2 n^1 = 0, \\
&(K^2) && - (n^2 + \gamma^{21} + \gamma^{23})(u'_2 - \psi'_2)/k^2_s + \gamma^{32}(\hat{u}'_{32} - \hat{\psi}'_{32})/\hat{k}^{32}_s \\
& && + \gamma^{42}(\hat{u}'_{42} - \hat{\psi}'_{42})/\hat{k}^{42}_s + \lambda_2 n^2 = 0, \\
&(K^3) && - (n^3 + \gamma^{31} + \gamma^{32})(u'_3 - \psi'_3)/k^3_s + \gamma^{23}(\hat{u}'_{23} - \hat{\psi}'_{23})/\hat{k}^{23}_s \\
& && + \gamma^{43}(\hat{u}'_{43} - \hat{\psi}'_{43})/\hat{k}^{43}_s + \lambda_2 n^3 = 0, \\
&(B^1_1) && n^1 u'_1 - \gamma^{21}\hat{u}'_{21} - \gamma^{31}\hat{u}'_{31} - \lambda_1 n^1 = 0, \\
&(B^2_1) && (n^2 + \gamma^{21} + \gamma^{23})u'_2 - \gamma^{32}\hat{u}'_{32} - \gamma^{42}\hat{u}'_{42} - \lambda_1 n^2 = 0, \\
&(B^3_1) && n^3 + \gamma^{31} + \gamma^{32})u'_3 - \gamma^{23}\hat{u}'_{23} - \gamma^{43}\hat{u}'_{43} - \lambda_1 n^3 = 0, \\
&(B^1_2) && n^1\psi'_1 - \gamma^{21}\hat{\psi}'_{21} - \gamma^{31}\hat{\psi}'_{31} - \lambda_2 n^1 = 0,
\end{aligned}$$

$$(B_2^2) \qquad (n^2 + \gamma^{21} + \gamma^{23})\psi_2' - \gamma^{32}\hat{\psi}_{32}' - \gamma^{42}\hat{\psi}_{42}' - \lambda_2 n^2 = 0,$$
$$(B_2^3) \qquad (n^3 + \gamma^{31} + \gamma^{32})\psi_3' - \gamma^{23}\hat{\psi}_{23}' - \gamma^{43}\hat{\psi}_{43}' - \lambda_2 n^3 = 0.$$

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

$$\begin{aligned}
MRS_c^1 =&(1+k_s^1) - \frac{\gamma^{21}\hat{\psi}_{21}'}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{21}) + (1 - k_s^1/\hat{k}_s^{21})(M\hat{R}S_c^{21} - 1)\right] \\
&- \frac{\gamma^{31}\hat{\psi}_{31}'}{n^1\lambda_2}\left[(MRS_c^1 - M\hat{R}S_c^{31}) - (1 - k_s^1/\hat{k}_s^{31})(M\hat{R}S_c^{31} - 1)\right] < 1 + k_s^1, \\
MRS_c^2 =&(1+k_s^2) - \frac{\gamma^{32}\hat{\psi}_{32}'}{n^2\lambda_2}\left[(MRS_c^2 - M\hat{R}S_c^{32}) - (1 - k_s^2/\hat{k}_s^{32})(M\hat{R}S_c^{32} - 1)\right] \\
&- \frac{\gamma^{42}\hat{\psi}_{42}'}{n^2\lambda_2}\left[(MRS_c^2 - M\hat{R}S_c^{42}) - (1 - k_s^2/\hat{k}_s^{42})(M\hat{R}S_c^{42} - 1)\right] \\
&< 1 + k_s^1 \text{ if } \gamma^{32} = 0, \\
MRS_c^3 =&(1+k_s^3) - \frac{\gamma^{23}\hat{\psi}_{23}'}{n^3\lambda_2}\left[(MRS_c^3 - M\hat{R}S_c^{23}) - (1 - k_s^3/\hat{k}_s^{23})(M\hat{R}S_c^{23} - 1)\right] \\
&- \frac{\gamma^{43}\hat{\psi}_{43}'}{n^3\lambda_2}\left[(MRS_c^3 - M\hat{R}S_c^{43}) - (1 - k_s^3/\hat{k}_s^{43})(M\hat{R}S_c^{43} - 1)\right] \\
&< 1 + k_s^1 \text{ if } \gamma^{23} = 0, \\
MRS_c^4 =&\frac{\lambda_1}{\lambda_2} = 1 + k_s^4.
\end{aligned}$$

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. Let's look at the case where $\gamma^{32} = 0$ and $\gamma^{23} > 0$ (both could though well be zero). Since both $\frac{\partial s}{\partial w} > 0$ and $\frac{\partial s}{\partial a} > 0$, the inequality in $\hat{s}^{23} \lesseqgtr s^3$ cannot be determined and therefore the inequality in $\frac{u_3'}{\psi_3'} \lesseqgtr \frac{\hat{u}_{43}'}{\hat{\psi}_{43}'}$ is not known as well as the inequality in $k_s^3 \lesseqgtr \hat{k}_s^{23}$ (this can be done by performing the same analysis as is done in appendix A). Hence, it cannot proven whether $\frac{u_3'}{\psi_3'}$ is smaller or bigger that $1 + k_s^3$. The same applies to the case when $\gamma^{32} > 0$ and $\gamma^{23} = 0$.

D Calculations for section 6.1

Case 1: $k(s, m)$

When the government observes Y, s and $K = k(s, m) - m$, individuals maximize utility w.r.t. m subject to the constraint $k(s, m) - m = K$, this implicitly defines $m(s, K)$. This means that the expenditure on m simply depends on the budget chosen and not on the type. I.e., 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 $K = k(s, m, a) - m$, individuals maximize utility w.r.t. m subject to the constraint $k(s, m, a) - m = K$. This means that individuals simply choose the m that will give them the K reported by the government. By the constraint $k(s, m, a) - m = K$, I implicitly define $m(s, K, a)$, with

$$\begin{aligned} m_w &= 0, \\ m_a &= -\frac{k_a}{k_m - 1} < 0. \end{aligned}$$

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, also I use the fact that $\frac{d\hat{m}^{21}}{dm^1} = \frac{k_m^1 - 1}{\hat{k}_m^{21} - 1} < 1$, since $\frac{dk_m}{da} = k_{mm} m_a + k_{ma} > 0$. The Lagrangian for this problem is

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

where $U^i = u(B_1^i - s^i - m^i) + \psi(B_2^i + s^i + m^i) + v(1 - Y^i/w)$ and $\hat{U}^{21} = u(B_1^1 - s^1 -$

$m^1) + \psi(B_2^1 + s^1 + \hat{m}^{21}) + v(1 - Y^1/w)$. The necessary conditions for type 1 are

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

Manipulation similar to those performed in appendix A will lead to (19).

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. In solving the model, it is simpler to set the bundles in terms of E^i, which gives a certain value of K^i, i.e. such that $k(E^i, s^i, m^i) - m^i = K^i$. Individuals choose both E and m in order to receive a certain value of K and the government has to take this into account when choosing E. From the individuals point of view, they freely choose E and m and face the following constraint: $k(E, s, m) - m = K$. This implicitly defines $E(K, s, m)$. Deriving partial derivatives is done by differentiating $k(E(K, s, m), s, m) - m - K = 0$. Then the individual's problem and necessary condition are, respectively

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

where $E_m = -(k_m - 1)/k_E < 0$, where $k_m > 1$. Performing comparative statics on the necessary condition yields

$$
\begin{aligned}
\frac{dm}{dw} &= \frac{v''Yw^{-2}E_m}{u'' + \psi'' - v'E_{mm} + v''E_m^2} < 0, \\
\frac{dm}{da} &= \frac{v'E_{ma} - v''E_m E_a}{u'' + \psi'' - v'E_{mm} + v''E_m^2} < 0,
\end{aligned}
$$

where $E_{mm} = [(k_{EE}E_m + k_{Em})(k_m - 1) - (k_{mm} + k_{mE}E_m)k_E]/k_E^2 > 0$, $E_{ma} = [(k_{Ea} + k_{EE}E_a)(k_m - 1) - k_{mE}E_a k_E]/k_E^2 > 0$ and $E_a = -k_a/k_E < 0$. This means that mimickers will spend less money on financial advisors and will therefore have a higher first period income than the type being mimicked, while they will have the same second period consumption.

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

$$\frac{dE}{da} = E_m m_a + E_a,$$

$$\frac{dE}{da} U_{mm} = (u'' + \psi'')E_a + v'[E_{ma}E_m - E_{mm}E_a]$$

$$= (u'' + \psi'')E_a + v' k_E^{-2}[(k_{Ea}E_m - k_{Em}E_a)(k_m - 1) + k_{mm}k_E E_a]. \quad \text{(D.1)}$$

where derivatives of $m(\cdot)$ follow from the necessary condition for m and derivatives of $E(\cdot)$ follow from the constraint $k(E, s, m) - m = K$. The bracket in (D.1) is positive if $k_{Em} > 0$, since then $k_{Ea}E_m - k_{Em}E_a = 0$ following the weakly separable form assumed. This means that when $k_{Em} > 0$, then $\frac{dE}{da} < 0$.

The bracket (D.1) in the last expression has an ambiguous sign. If $k(E, s, m, a)$ takes the following weakly separable form, $f_1(E)f_2(s)f_3(m)g(a)$ with $f_3(m) = m^\alpha$, then the bracket will be positive if $\alpha < \frac{1}{2}$.[18] I believe this assumption to be quite reasonable. If this is the case, then $\frac{dE}{da} < 0$. The above mentioned assumption is also only a sufficient condition, not a necessary condition for $\frac{dE}{da} < 0$. I therefore conclude that $\frac{dE}{da} < 0$. This means that mimickers will have more leisure, both when they have higher a and w.

The Lagrangian for the government's problem is

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

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

$$(E^1) \qquad -n^1 v_1' + \gamma \hat{v}_{21}' \frac{k_E^1}{\hat{k}_E^{21}} + \lambda_2 n^1 k_E^1 = 0,$$

[18]When $k(E, s, m, a) = f_1(E)f_2(s)m^\alpha g(a)$, then

$$k_{Ea}E_m k_m + k_{mm}k_E E_a = -k_E^{-1}[k_{Ea}k_m^2 + k_{mm}k_E k_a] = -k_E^{-1} f_1' g' (f_1 f_2 g)^2 f_2 f_3 \alpha m^{2\alpha-2}[\alpha - (1-\alpha)]$$

which is positive if $\alpha < \frac{1}{2}$, if this holds then the bracket will be positive since $k_{Ea}E_m < 0$.

$$(s^1) \qquad n^1(\psi_1' - u_1') - \gamma(\psi_1' - \hat{u}_{21}') + \lambda_2 n^1 k_s^1 = 0,$$

$$(B_2^1) \qquad (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0,$$

where $\frac{d\hat{E}^{21}}{dE^1} = \frac{k_E^1}{\hat{k}_E^{21}} < 1$ follows from $k(E^1, m^1, s^1, a^1) - m^1 - k(\hat{E}^{21}, \hat{m}^{21}, s^1, a^2) + \hat{m}^{21} = 0$ and $\frac{dk_E}{da} = k_{Ea} + k_{EE}\frac{dE}{da} > 0$.

A different way of approaching is problem is to let the government choose m instead of E, the necessary condition for m^1 is

$$(m^1) \qquad n^1(\psi_1' - u_1') - \gamma(\psi_1' - \hat{u}_{21}') + \lambda_2 n^1(k_m^1 - 1) = 0.$$

This shows that the government will set $k_s^1 = k_m^1 - 1$. In other words, the optimal choice between s and m will not be distorted. From (20a) and noting that $-(u' - \psi') = -v'E_m = v'(k_m - 1)/k_E$, I get

$$\begin{aligned}
k_s T_K =& \frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2}\left\{(MRS_c^1 - 1) - (M\hat{R}S_c^{21} - 1)\frac{k_m^1 - 1}{\hat{k}_m^{21} - 1}\right\} \\
=& \frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2}\left\{(MRS_c^1 - M\hat{R}S_c^{21}) + (M\hat{R}S_c^{21} - 1)\left(1 - \frac{k_m^1 - 1}{\hat{k}_m^{21} - 1}\right)\right\}.
\end{aligned}$$

Subtracting this from $k_s T_K + T_s = \frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2}\left\{MRS_c^1 - M\hat{R}S_c^{21}\right\}$, which follows from (20b), I get

$$T_s = -\frac{\gamma\hat{\psi}_{21}'}{n^1\lambda_2}(M\hat{R}S_c^{21} - 1)\left(1 - \frac{k_m^1 - 1}{\hat{k}_m^{21} - 1}\right) < 0,$$

where $k_m^1 < \hat{k}_m^{21}$, since $\frac{dk_m}{da} = k_{mm}m_a + k_{ma} > 0$.

E Calculations for section 6.2

The government observes Y^i, s^i and K^i, and offers bundles in terms of these variables for both types of individuals. In solving the model, it well be simpler to set the bundles in terms of E^i, which gives a certain value of K^i, i.e. such that $k(E^i, s^i, m^i) - m^i = K^i$. Individuals choose both E and m in order to receive a certain value of K and the government has to take this into account when choosing E^i. From the individual point of view, he freely chooses E and m and faces the following constraint: $k(E, s, m) - m = K$. This implicitly defines $E(K, s, m)$. Deriving partial derivatives is done by differentiating $k(E(K, s, m), s, m) - m - K = 0$. The individual's problem and necessary condition are, respectively

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

where $E_m = -(k_m - 1)/k_E$. This defines $m(Y, B_1, s, K, e)$. Performing comparative statics on the necessary condition yields

$$\frac{dm}{de} = \frac{u''}{u'' + \psi'' + v'' E_m^2 - v' E_{mm}} \in (0, 1),$$
$$\frac{dE}{dm} = E_m m_e < 0,$$

where $E_{mm} = (k_{EE} E_m (k_m - 1) - k_{mm} k_E)/k_E^2 > 0$. This implies that mimickers have a higher (lower) first (second) period consumption than the type being mimicked, since $e^1 - m^1 < e^2 - \hat{m}^{21}$. Due to the constraint $k(E, m, s) - m = K$, mimickers will have lower investment effort. This means that $u_1' > \hat{u}_{21}'$, $\psi_1' < \hat{\psi}_{21}'$, $v_1' > \hat{v}_{21}'$ and $k_E^1 < k_E^{21}$.[19] Thereby $MRS_c^1 = \frac{u_1'}{\psi_1'} > \frac{\hat{u}_{21}'}{\hat{\psi}_{21}'} = M\hat{R}S_c^{21}$ and $MRS_K^1 = \frac{v_1'}{\psi_1'} \frac{1}{k_E^1} > \frac{v_1'}{\psi_1'} \frac{1}{\hat{k}_E^{21}} = M\hat{R}S_K^{21}$.

The Lagrangian for the government's problem is

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

where $U^i = u(B_1^i + e^i - m^i) + \psi(B_2^i) + v(1 - Y^i/w - E^i)$ and $\hat{U}^{21} = u(B_1^1 + e^2 - \hat{m}^{21}) + \psi(B_2^1) + v(1 - Y^1/w - \hat{E}^{21}(E^1))$ and $\frac{d\hat{E}^{21}}{dE^1} = \frac{k_E^1}{\hat{k}_E^{21}} < 1$. The necessary conditions for type 1 are

$$(E^1) \qquad -n^1 v_1' + \gamma \hat{v}_{21}' \frac{k_E^1}{\hat{k}_E^{21}} + \lambda_2 n^1 k_E^1 = 0,$$
$$(Y^1) \qquad -n^1 v_1' \frac{1}{w} + \gamma \hat{v}_{21}' \frac{1}{w} + \lambda_1 n^1 = 0,$$
$$(s^1) \qquad n^1 (\psi_1' - u_1') - \gamma (\psi_1' - \hat{u}_{21}') + \lambda_2 n^1 k_s^1 = 0,$$
$$(B_2^1) \qquad (n^1 - \gamma) \psi_1' - \lambda_2 n^1 = 0.$$

Performing an identical manipulation as the one in the end of appendix D shows that

[19] This can be shown formally by $\frac{dk_E}{de} = k_{Em} m_e + k_{EE} E_m m_e > 0$.

T_s is positive,

$$T_s = -\frac{\gamma \hat{\psi}'_{21}}{n^1 \lambda_2}(M\hat{R}S_c^{21} - 1)\left(1 - \frac{k_m^1 - 1}{\hat{k}_m^{21} - 1}\right) > 0,$$

where $k_m^1 > \hat{k}_m^{21}$, since $\frac{dk_m}{de} = k_{mm}m_e < 0$. Also, performing manipulations identical to those in appendix A, will show the the optimal allocation resulting from this model take the same form as in (20a) and (20b), with the same inequalities.

Difference in distortion between Y and K In order to analyst whether the marginal tax rate on labour income is larger or smaller than the marginal tax rate on capital income, I follow the same procedure as in section 5.2 and consider a model where labour income and capital income are both received in the second period. Otherwise, the setup is unchanged. In the absence of the government, the individual budget constraint is

$$c_1 = e - s - m,$$
$$c_2 = Y + s + k(E, s, m).$$

The Lagrangian for the government's problem is

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

The necessary conditions for type 1 are

$$(E^1) \qquad -n^1 v_1' + \gamma \hat{v}'_{21}\frac{k_E^1}{\hat{k}_E^{21}} + \lambda_2 n^1 k_E^1 = 0,$$
$$(Y^1) \qquad -n^1 v_1' \frac{1}{w} + \gamma \hat{v}'_{21}\frac{1}{w} + \lambda_2 n^1 = 0,$$
$$(B_2^1) \qquad (n^1 - \gamma)\psi_1' - \lambda_2 n^1 = 0.$$

This can be manipulated to get the following conditions

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

Also, the relationship between MRS_K and MRS_Y can be written as

$$MRS_Y = MRS_K \frac{k_E}{w}, \tag{E.1}$$

which holds both for both types as well as a mimicker. Multiplying the necessary condition for Y^1 with w and rearranging, it follows that

$$\begin{aligned} n^1 v_1' - \gamma \hat{v}_{21}' \frac{k_E^1}{\hat{k}_E^{21}} =& \lambda_2 n^1 k_E^1 \\ n^1 v_1' - \gamma \hat{v}_{21}' =& \lambda_2 n^1 w \end{aligned}$$

Since $\frac{k_E^1}{\hat{k}_E^{21}} < 1$, it follows that at the optimum, $k_E^1 > w$. From (E.1) it follows that $MRS_Y^1 > MRS_K^1$.

I introduce the budget constraint in the second period with a tax function as $c_2 = Y + s + k(E, s, m,) - T(Y, k(E, s, m))$. The individual necessary conditions for Y and E are $MRS_Y^1 = 1 - T_Y$ and $MRS_K^1 = 1 - T_K$, respectively. Since $MRS_Y^1 > MRS_K^1$, it follows that $T_K > T_Y$.

F 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 (22a) and (22b).

Demand for b The individual problem can be presented similar to (10)

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

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 analyst 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 $b_{B_2} = b_{B_1} \frac{dB_1}{dB_2}\Big|_{\overline{U}} = -b_{B_1}\frac{\psi'}{u'}$. To sign the compensated effect of r, I note that $\frac{\partial b}{\partial B_2}\frac{dB_2}{dB_1}\Big|_{\overline{U}} = \frac{\partial b}{\partial B_1}\frac{dB_2}{dB_2}\Big|_{\overline{U}} = -\frac{\partial b}{\partial B_1}$ since $\frac{dB_2}{dB_2}\Big|_{\overline{U}} = -1$ and that

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

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

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

Behavior 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 analyst the behavior 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

$$\begin{aligned} U_{ss} &= u'' + \psi'' - v'E_{ss} + v''E_s^2 < 0, \\ U_{bb} &= u'' + \psi''(1+r)^2 + v''E_b^2 < 0, \\ U_{sb} &= -u'' - \psi''(1+r) + v''E_sE_b > 0, \end{aligned}$$

$$
\begin{aligned}
U_{sa} &= -v' E_{sa} + v'' E_s E_a < 0, \\
U_{ba} &= -v' E_{ba} + v'' E_b E_a < 0, \\
E_{ss} &= [k_{EE} + 2k_{Es} k_s - k_{ss} k_E]/k_E^2 > 0, \\
E_{bb} &= 0, \\
E_{sa} &= [k_{EE} E_a k_s - k_{sE} E_a k_E]/k_E^2 > 0, \\
E_{ba} &= -[r k_{EE} E_a + r k_{Ea}]/k_E^2 < 0.
\end{aligned}
$$

The effects of an increase in a on savings is negative

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

where I made use of the following

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

The effects of an increase in a on b is ambiguous

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

A mimicker will have a lower first period consumption,

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

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

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

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