tinbergen institute

Alexandra Victoria Rusu

Essays in Public Economics

Erasmus University Rotterdam

Essays in Public Economics

ISBN: 978 90 3610 600 9

This book is no. 760 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays in Public Economics

Essays over de economie van de publieke sector

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Friday 27 March 2020 at 11:00 hrs

by

ALEXANDRA VICTORIA RUSU

born in Iasi, Romania

Erasmus University Rotterdam

Doctorate Committee

Promotors:	Prof. dr. B. Jacobs
	Prof. dr. H.D. Webbink
Other members:	Prof. dr. A. L. Bovenberg
	Prof. dr. R. Dur
	Prof. dr. C. van Ewijk
	Prof. dr. A. M. Lejour
	Prof. dr. D. S. Schindler

Contents

Chapter 1

Introduction

The present dissertation consists of two distinct parts: Chapters 2 and 3 study how governments should set taxes in order to raise funds efficiently, while Chapter 4 analyzes how public funds are spent. Despite employing different methods and covering disparate topics, these two parts accurately reflect my research agenda: understanding how to make government finances more efficient. Through my own experience of growing up in a transition country, I have learned that how the government spends its funds is as important as how it obtains it: assuming taxes are set optimally, public funds can be squandered through corrupt or inefficient practices. Alternatively, inefficient taxation can lead to insufficient funds in the government coffers, decreasing its ability to provide essential services.

The first part consists of Chapters 2 and 3 and covers a question that is currently high on the research agenda of both academics and policymakers: whether and how to tax capital income.

Chapter 2 offers an intuitive explanation for the Chamley-Judd result, a well-known result stating capital income should not be taxed in the steady state (Chamley, 1986; Judd, 1985). In the steady state, consumption demands in each period become equally complementary to leisure. This makes taxes on capital income redundant: they cannot alleviate distortions from taxing labor income, but they do distort intertemporal consumption decisions.

The explanation is rooted in Corlett and Hague (1953): if goods that are stronger complements to leisure are taxed relatively more, individuals substitute away from leisure by working more. To use an example familiar to many readers, such an argument could be used to advocate for subsidizing lunchtime alcohol in France, as it is complementary to labor due to business lunches.[1] We show that the steady-state assumption makes present and future consumption equally complementary to labor, regardless of the type of utility function. Thus, a differentiated consumption tax (or a capital

[1] A similar argument is used for childcare subsidies, which are arguably less contentious: if childcare is strongly complementary to labor (Jaumotte, 2003), it should be subsidized to increase labor supply and alleviate distortions caused by labor taxes.

income tax) would not bring any benefits on top of a labor income tax, while distorting consumption decisions.

The explanation presented in Chapter 2 bridges the macroeconomics and commodity tax literatures, showing how the intuition for the Chamley-Judd result is consistent with standard complementarity arguments in public economics. The Chapter also shows how two explanations previously offered by the literature can be misleading. The first argues that capital income taxes can never be optimal become they impose an ever-growing tax burden on future consumption (Judd, 1999; Banks and Diamond, 2010). This logic is only applicable when strong restrictions are made on the utility function, ensuring that the Ramsey tax smoothing intuition is equivalent to the more general Corlett-Hague logic. The second explanation argues that in the steady state, all taxes on capital income are shifted to labor due to general-equilibrium effects on factor prices, so labor taxes should be the instrument of choice (Auerbach and Kotlikoff, 1983; Correia, 1996; Mankiw, Weinzierl, and Yagan, 2011; Piketty and Zucman, 2013). The Chapter shows that this is not a necessary condition: it shuts off any general-equilibrium effects by studying a model with exogenous factor prices and still obtains that capital income taxes should be zero in the steady state. Thus, general equilibrium effects cannot be the explanation for the Chamley-Judd result, as they are absent in the partial-equilibrium model.

The Chamley-Judd result has been a cornerstone of public economics for more than 30 years, with economists strongly arguing against capital income taxation on its basis. However, the results in Chapter 2, together with the analysis of Straub and Werning (2020) about the existence of steady states, suggest that the case for zero taxation of capital incomes is not as clear-cut as previously thought. Since the result relies on the assumption of a steady-state or of a specific type of utility function, the Chamley-Judd result is merely a technical result occurring under very specific conditions, rather than a general result which should be informative for policy.

Chapter 3 aims to adapt optimal-tax models to fit the real world by exploring how capital incomes should be taxed when individuals have heterogeneous returns to capital. With mounting empirical evidence showing large differences in rates of return to capital across percentiles of the wealth distribution (Fagereng et al., 2020; Campbell, Ramadorai, and Ranish, 2018), this question becomes increasingly relevant from a policy perspective.

The Chapter generalises a two-period version of the Mirrlees (1971) model to include returns to capital that can vary both with individual ability and with savings. The model embeds multiple microfoundations for heterogeneity in capital returns and enables the study of their effects on optimal tax policy when the government has access to fully non-linear tax schedules. The focus is on two cases: the first features high-ability individuals having access to closely-held investments that generate excess returns, while the second features increasing scale returns due to rich individuals having stronger incentives to invest in financial knowledge and advice.

Chapter 3 demonstrates that it is optimal to positively tax capital incomes in both cases. When capital incomes are positively correlated with ability, they reveal information about ability over and above the labor income base. Since, in a second-best world, governments aim to spread distortions among all bases which reveal information about ability, capital income becomes a natural base for taxation. Conversely, even when capital incomes do not directly correlate with ability but are an increasing function of savings, it is optimal to tax capital incomes. In this case, thanks to increasing scale returns, rich individuals can obtain higher returns to capital than poorer ones. This makes it optimal for the government to redistribute later in life, once the rich individuals have realised large capital returns: rich individuals strongly prefer paying taxes later in life, while poor individuals have small capital returns, making them relatively indifferent between early and late redistribution.

In addition to showing that capital income taxes are positive in a wide range of cases, Chapter 3 also obtains simplified expressions for optimal taxes in terms of empirically-measurable elasticities and characteristics of the capital and labor income distributions. Furthermore, the numerical simulations calibrated to the US case show two important features of optimal capital income taxes when heterogeneity in returns to capital is due to closely-held assets:

- Marginal capital income taxes are economically significant in all cases.
- Marginal income taxes increase for most of the income distribution.

Chapter 4 forms the second part of the present thesis and studies possible instance of influence peddling and conflict of interest in Romanian public hospitals. I study the connections between pharmaceutical companies sponsoring doctors in public hospitals and the procurement contracts those hospitals sign with various pharmaceutical firms. Sponsoring a doctor with management responsibilities is more strongly correlated with the probability of a direct contract (i.e, without tender) occurring than sponsoring a regular doctor, but the difference is not economically significant for contracts awarded with tenders. I document a timing effect: within three months of a sponsorship, there is an increase in the probability of a procurement contract occurring between a sponsored hospital and a sponsoring firm. Furthermore, procurement contracts linked to sponsorships are larger than those not linked. Together with the institutional environment and evidence suggesting contracts linked to sponsorships are less transparent, this evidence can be interpreted more in line with sponsorships acting as kickbacks, rather than legitimate marketing means.

Thus, the main message of this Chapter is that there are reasons to believe the doctors who accept sponsorships from pharmaceutical companies are in a situation of conflict of interest. While an outright ban of the practice of pharmaceutical sponsorships would probably hurt the public healthcare system by canceling one of the few sources of financial support for doctors' continuous education, I believe that the current system needs to be reformed in order to cut the direct link between companies and doctors.

Chapter 2

Why is the Long-Run Tax on Capital Income Zero? Explaining the Chamley-Judd Result*

*This Chapter is based on Jacobs and Rusu (2018). The authors would like to thank Robin Boadway, Emmanuel Saez and seminar participants at the Tinbergen Institute Rotterdam for useful comments and suggestions. All remaining errors are our own.

2.1 Introduction

Should capital income be taxed or not? This is one of the oldest and most important questions in public finance. However, the literature has not yet settled on a definite answer and the issue remains controversial from a policy perspective.[1] The arguments against taxing capital income rely on Chamley (1986) and Judd (1985), who suggest that in the long run the required revenue should be generated solely through taxing labour income. Thus, it is never optimal to tax capital income in the long run, but it might be optimal to tax it in the short run. Although there is a large literature on the robustness of the zero-capital income tax result,[2] the economic mechanism and the intuition for the zero tax result remain elusive.[3]

[1] For example, the main editors of the Mirrlees Review conclude that taxing the (normal) return to savings is undesirable (Mirrlees et al., 2011) However, Banks and Diamond (2010), who also write a chapter in the Mirrlees Review, argue that taxing the returns to capital is optimal. Mankiw, Weinzierl, and Yagan (2011) argue against taxing capital income in the Journal of Economic Literature, whereas Diamond and Saez (2011) argue in favor of taxing capital income in that very same journal.

[2] See e.g. Jones, Manuelli, and Rossi (1997), Krusell, Kuruşçu, and Smith (2010) and Straub and Werning (2020).

[3] Erosa and Gervais (2002) use an OLG version of the Ramsey models in Chamley (1986) and Judd (1985) to demonstrate that the optimal tax on capital income is generally non-zero.

In this paper, we argue that the zero capital income tax result can be explained with standard principles from the theory of optimal commodity taxation. The tax on capital income should be seen as a differentiated tax on consumption at different dates, so that in the optimum, it should be zero if optimal consumption taxes are uniform. The main intuition for optimal uniform commodity taxation in the Ramsey (1927) framework is found in Corlett and Hague (1953): if goods that are stronger complements to leisure are taxed at higher rates, individuals substitute away from leisure and work more. Since labour supply is distorted downwards, commodity tax differentiation can alleviate distortions of the labour income tax, but at the expense of distorting commodity demands. Formally, uniform commodity taxation is optimal if the utility function is weakly separable between consumption and leisure and homothetic in consumption (Sandmo, 1974).[4,5] In that case, different commodities are equally complementary to leisure and commodity tax differentiation only causes goods market distortions, without alleviating labour market distortions.

We analyze a version of the Chamley-Judd model due to Ljungqvist and Sargent (2004), which is closely related to Chamley (1986) and Judd (1999).[6] An infinitely-lived representative agent decides how much to work and save in each period. The government needs to finance an exogenous stream of outlays and optimises linear taxes on labour and capital income such that the lifetime utility of the representative individual is maximised. To avoid a degenerate steady state or a first-best solution, we assume that initial capital endowments are null and first-period production only uses labour.[7] This assumption also avoids problems with incomplete tax codes (Judd, 1985; Correia, 1996; Abel, 2007; Chari, Nicolini, and Teles, 2018). In the steady state, optimal taxes on capital income are shown to be zero. Our explanation for this result is that the steady-state assumption in Chamley (1986) and Judd (1999) forces consumption in each period to become equally complementary to leisure at all times. Proportional taxes on capital income impose the same distortions on labour supply as proportional taxes on labour income, but in addition also distort saving. Therefore, the

[4]Deaton (1979) demonstrates that uniform commodity taxation is even obtained in settings with heterogeneous agents if preferences are of the Gorman (1961) polar form, resulting in quasi-homothetic preferences. However, uniform commodity taxation can then only be obtained if the government has access to a (non-individualized) lump-sum tax. This instrument is ruled out in the Chamley-Judd setting with a representative agent to obtain a non-trivial second-best analysis.

[5]The Corrlett-Hague motive for differentiated commodity carries over to Mirrleesian frameworks with optimal non-linear taxation of labour income, cf. Atkinson and Stiglitz (1976) and Jacobs and Boadway (2014). The Atkinson and Stiglitz (1976) theorem shows that uniform commodity taxation is optimal if the government can levy a non-linear tax on labour income and preferences are weakly separable between consumption over time and leisure. Hence, quasi-homotheticity is no longer required to obtain optimally uniform commodity taxation if income taxes are non-linear. See also Ordover and Phelps (1979) for an application of optimal taxes on capital income in a 2-period OLG framework with optimal non-linear taxes on labour income.

[6]There are two reasons for doing so. First, the model we use is the most common formulation in the literature and is presented as the workhorse argument in standard macroeconomics curricula. Second, as Straub and Werning (2020) and Lansing (1999) have shown, the results in the two-type model of Judd (1985) are very sensitive to model assumptions, e.g. they depend crucially on the value of the intertemporal elasticity of substitution.

[7]Straub and Werning (2020) showed that the size of initial government debt can determine the existence and nature of the steady-state, which are issues that we want to avoid.

government should not distort intertemporal consumption decisions in order to alleviate labour supply distortions and optimal taxes on capital income should become zero.[8] We thus establish a close link between the zero-tax result in Chamley (1986) and Judd (1999) and the theory of optimal commodity taxation. A similar point was made by Stiglitz (2018), who reflected on the implications of the original Atkinson-Stiglitz theorem for capital income taxation in a model with heterogeneous agents. Although his conclusions are broadly similar, Stiglitz (2018) focused on how taxation can be used to soften incentive compatibility constraints. We focus on a standard macro model with a representative agent and show how standard macro assumptions link to the more micro- results in models such as that of Stiglitz (2018).

By showing that standard optimal taxation principles underlie the zero tax on capital income, we reveal that the explanations previously offered by the literature can be misleading. The first intuition, provided by Judd (1999) and subsequently used in Banks and Diamond (2010), argues that the economy need not converge to a steady state for the optimal long-run tax on capital income to be zero. Since capital income taxes impose an exponentially-growing tax burden on consumption in the more distant future, it can never be optimal to set them to strictly positive rates in the long run. Such an explosive path of tax distortions in finite time is incompatible with standard Ramsey principles, which insist that tax distortions be smoothed out over time. Therefore, in order to rule out exponentially growing tax burdens, taxes on capital income should become zero in finite time. We agree with Judd (1999) that the intuition for Chamley-Judd result should be firmly rooted in optimal taxation principles. However, the Ramsey logic is applicable only when consumption demands depend solely on own prices. Hence, strong restrictions need to be made on the utility function: additive separability over time and separability between consumption and leisure. Only under these restrictions is the Ramsey tax smoothing intuition equivalent to the more general Corlett-Hague logic; the commodities that are less price elastic are also the commodities that more complementary to leisure. See also Atkinson and Stiglitz (1980, Ch. 12).

Judd (1999) argued that convergence to a steady-state is not required in order to get zero optimal capital income taxes. In finite time, capital income taxes are zero either if the multipliers on the government budget constraints are bounded or if preferences are such that the multipliers are constant. Straub and Werning (2020) correctly criticise imposing constraints on endogenous multipliers, since doing so boils down to <u>assuming</u> that the optimal tax on capital income is zero. We add to the analysis in Straub and Werning (2020) by showing that the multipliers on the government budget constraints

[8] Our paper is meant as a positive, methodological contribution aimed at clarifying the result of Chamley (1986) and Judd (1999) that capital income should not be taxed in the long run. It is not meant to serve as a normative policy prescription. Banks and Diamond (2010), Diamond and Saez (2011) and Jacobs (2013) have argued that capital income should be taxed at positive rates for various reasons that the framework of Chamley (1986) and Judd (1999) cannot address.

are constant only if preferences are such that consumption is equally complementary to leisure at all times and the optimal capital income tax is in fact zero in every period.

The second argument why capital income taxes are optimally zero can be found in Auerbach and Kotlikoff (1983), Correia (1996) and Mankiw, Weinzierl, and Yagan (2011). It is argued that in the steady state, all taxes on capital income are shifted to labour due to general-equilibrium effects on factor prices. Therefore, it is better to tax labour income directly and avoid distortions in the capital market. This argument relies on the notion that in the steady state, the net return to capital is completely determined by exogenous factors such as the depreciation rate and the rate of time preference. Consequently, any tax on capital income has to result in a one-to-one increase of the gross return to capital to keep the net return to capital constant. This requires a fall in the steady-state capital stock, which decreases wages. As a result, the tax burden is completely shifted to labour. We analyse an open-economy version of the Chamley-Judd model, where we switch off any general-equilibrium effects on factor prices that occur due to the taxation of capital income. This allows us to confirm the results of Diamond and Mirrlees (1971a) in a dynamic setting: the expressions for optimal taxes in partial equilibrium are identical to those obtained in general equilibrium.[9] Therefore, general-equilibrium effects in factor prices shifting the entire tax burden towards labour cannot be an explanation why capital income should not be taxed in the long-run. This contrasts with the impressions that are given in Auerbach and Kotlikoff (1983), Correia (1996) and Mankiw, Weinzierl, and Yagan (2011).

To our knowledge, our paper is the first contribution that binds together all explanations for the Chamley-Judd zero tax result through a single mechanism. In particular, our interpretation holds both inside and outside steady-state and in general- and partial- equilibrium settings. Furthermore, our interpretation is consistent both with the macroeconomics literature on capital income taxation and with the optimal taxation literature on commodity tax differentiation.

The present work complements the analysis of Straub and Werning (2020), who showed that the results in Chamley (1986) and Judd (1985) are not as general as previously thought. By using a model with additively separable time preferences, we focus on the only case identified by Straub and Werning (2020) where the capital stock is positive and taxes on capital income are zero in the steady state. Straub and Werning (2020) show that if preferences are not additively separable over time, the zero capital income tax in Chamley (1986) is imposed on a zero tax base, or it coexists with a zero labour income tax. We also explore a version of the model where preferences are not time-separable

[9]Judd (1999) argues that the zero capital tax result is also an application of the Diamond-Mirrlees production efficiency theorem. He claims that it is not optimal to tax capital, since it is an intermediate good. Diamond and Saez (2011, p.177, footnote 15) (correctly) argue that this interpretation is not applicable, since production is always efficient in the Chamley-Judd model in the absence of taxes at the firm level.

to show how the complementarity between consumption and leisure determines optimal taxes on capital income with non-additive preferences. In particular, we show that if preferences are not time-separable, but weakly separable between consumption and leisure and homothetic with respect to both, consumption at different times is equally complementary to leisure at all times. Consequently, the tax on capital income is (always) zero.

The rest of the paper is structured as follows. In the next section, we introduce the general-equilibrium model and show that the Corrlett-Hague motive for commodity tax differentiation vanishes in the steady state of the Chamley-Judd model. The reason is that consumption becomes equally complementary to leisure at all times. In the third section, we show how our interpretation relates to the other intuitions in the literature. A final section concludes. Proofs not covered in the main text can be found in the Appendix.

2.2 Long-run taxes on capital income in general equilibrium

2.2.1 Representative individual

This section starts with a general-equilibrium formulation of a closed economy as in Chamley (1986) and Judd (1999), where utility is time-separable and time is indexed by t. We follow the representation given in Ljungqvist and Sargent (2004). There is an infinitely-lived representative individual who maximises the discounted value of lifetime utility:

$$\sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \quad u_c, -u_l > 0, \quad u_{cc}, u_{ll} < 0. \tag{2.1}$$

The utility function $u(c_t, l_t)$ in each period is increasing, strictly concave and twice differentiable in both consumption c_t and leisure $1 - l_t$. The individual's pure rate of time preference is captured by the discount factor β and her assets are denoted by a_t.

The representative individual owns no assets in period 0 ($a_0 = 0$). Consequently, there is no motive to tax pure rents from (initial) asset endowments. By assuming zero initial assets, we avoid the possibilities of a degenerate steady state or a first-best solution, see also Straub and Werning (2020). The individual is endowed with one unit of time per period, which must be divided between work and leisure. In each period, labour has to satisfy a time constraint: $0 \leq l_t \leq 1$. The gross interest rate is r_t and the gross wage rate is w_t. The government levies a proportional tax on capital income τ_t^K and a proportional tax on labour income τ_t^L in every period. Consequently, the individual's budget constraint is:

$$a_{t+1} = (1 + (1 - \tau_t^K) r_t) a_t + (1 - \tau_t^L) w_t l_t - c_t, \quad t \geq 0, \quad a_0 = 0, \tag{2.2}$$

$$\lim_{t \to \infty} \frac{a_{t+1}}{\prod_{s=1}^{t} (1 + (1 - \tau_s^K) r_s)} = 0. \tag{2.3}$$

Equation (2.3) says that the present discounted value of the individual's terminal assets should be 0, thus ruling out explosive asset paths (a no Ponzi-scheme condition). By iterating the individual's budget constraint forward and applying the transversality condition in equation (2.3), we obtain her lifetime budget constraint:

$$\sum_{t=0}^{\infty} \frac{c_t}{\prod_{s=1}^{t} (1 + (1 - \tau_s^K) r_s)} = \sum_{t=0}^{\infty} \frac{(1 - \tau_t^L) w_t l_t}{\prod_{s=1}^{t} (1 + (1 - \tau_s^K) r_s)}. \tag{2.4}$$

The representative individual's problem consists of choosing sequences of consumption $\{c_t\}_{t=0}^{\infty}$, labour supply $\{l_t\}_{t=0}^{\infty}$ and assets $\{a_{t+1}\}_{t=0}^{\infty}$ such that lifetime utility (2.1) is maximised subject to the budget constraint (2.4). Assuming an interior solution for l_t and denoting the multiplier on the period t budget constraint by $\beta^t \lambda_t$, we can obtain the first-order conditions that govern optimal labour supply and saving behaviour:

$$u_{c_t} = \lambda_t, \quad t \geq 0, \tag{2.5}$$

$$-u_{l_t} = \lambda_t (1 - \tau_t^L) w_t, \quad t \geq 0, \tag{2.6}$$

$$\frac{\lambda_t}{\beta \lambda_{t+1}} = 1 + (1 - \tau_{t+1}^K) r_{t+1}, \quad t > 0. \tag{2.7}$$

Equation (2.5) states that in the optimum, the marginal benefit of consuming one extra unit, u_{c_t}, should be equal to the marginal cost λ_t of doing so. Similarly, equation (2.6) shows that the individual should work until the marginal cost of sacrificing one extra unit of leisure, $-u_{l_t}$, is equal to the gain in utility due to having more income, $\lambda_t (1 - \tau_t^L) w_t$. The Euler equation (2.7) describes the optimal allocation of consumption across time: the individual should save until her increase in utility from consuming marginally more in the current period (λ_t) is the same as her discounted increase in utility from investing that consumption increment at market prices and consuming it the next period, $\beta (1 + (1 - \tau_{t+1}^K) r_{t+1}) \lambda_{t+1}$.

2.2.2 Government

The government's objective is to maximise the representative individual's utility, while satisfying an exogenous revenue requirement g_t in every period. Like Chamley (1986) and Judd (1999), we assume that the government can credibly commit to the policies it sets.[10] Furthermore, we assume that the government can verify aggregate capital and labour income, but has no access to lump-sum taxes. Thus, it can use proportional taxes τ_t^L on labour income and τ_t^K on capital income and issuance of debt d_{t+1} to raise revenue.

We assume that in period 0, the initial level of debt is null: $d_0 = 0$. Since this is a deterministic model without default, government bonds and private assets are perfect substitutes. Perfect arbitrage thus ensures that the interest rate on government bonds equals the interest rate r_t on other assets. Hence, the period-by-period government budget constraint reads as:

$$d_{t+1} = (1 + r_t)d_t + g_t - \tau_t^L w_t l_t - \tau_t^K r_t a_t, \quad t \geq 0, \quad d_0 = 0, \tag{2.8}$$

$$\lim_{t\to\infty} \frac{d_{t+1}}{\prod_{s=1}^{t}(1 + r_s)} = 0. \tag{2.9}$$

The government debt d_{t+1} also has to satisfy transversality condition (2.9) to rule out explosive paths for public debt.

2.2.3 Firms

There is a single representative firm that uses capital k_t and labour l_t to produce output. In all periods $t > 0$, the production function is given by $f(k_t, l_t)$, which exhibits constant returns to scale, satisfies the Inada conditions and features positive and decreasing marginal returns to both capital and labour: $f_l, f_k > 0$, $f_{ll}, f_{kk} < 0$.[11] Capital depreciates at rate δ. In period 0, the production function uses only labour: $f_0 = A_0 l_0$. This ensures that endowments are not required for starting the production process. Profit maximisation implies that marginal products equal marginal costs in each period:

$$f_k(k_t, l_t) = r_t + \delta, \quad t > 0, \tag{2.10}$$

$$f_l(k_t, l_t) = w_t, \quad t > 0, \tag{2.11}$$

$$A_0 = w_0. \tag{2.12}$$

There are no pure profits in each period due to constant returns to scale in production.

[10] There is a well-known time-consistency problem in the optimal setting of capital taxes. Once capital is accumulated, capital owners cannot respond by withdrawing their investment. Hence, the government has an incentive to expropriate individuals by levying a tax on capital to reduce distortionary labour taxes (Kydland and Prescott, 1977; Fischer, 1980).

[11] The Inada conditions are: $\lim_{k_t\to 0} f_k(k_t, l_t) = \lim_{l_t\to 0} f_l(k_t, l_t) = \infty$ and $\lim_{k_t\to\infty} f_k(k_t, l_t) = \lim_{l_t\to\infty} f_l(k_t, l_t) = 0$.

2.2.4 General equilibrium

Equilibrium in the goods market requires that the total demand for goods – private consumption c_t, public consumption g_t, investment $k_{t+1} - (1-\delta)k_t$ – equals the supply of goods:

$$c_t + g_t + k_{t+1} - (1-\delta)k_t = f(k_t, l_t), \quad t \geq 0. \tag{2.13}$$

Equilibrium in the capital market requires that the demand for capital by firms k_t and demand of government debt d_t equal the supply of assets by the representative individual a_t:[12]

$$k_t + d_t = a_t. \tag{2.14}$$

2.2.5 Primal approach in general equilibrium

The government's problem is to choose the sequence of taxes $\{\tau_{t+1}^K, \tau_t^L\}_{t=0}^{\infty}$ that maximises the representative individual's lifetime utility. In order to derive the optimal tax rules, we employ the primal approach to the optimal tax problem. First, the government optimally derives the second-best allocation $\{c_t, l_t, g_t, k_{t+1}\}_{t=0}^{\infty}$ subject to the resource and implementability constraints. Second, this allocation is decentralised using the tax instruments to obtain the same allocation as the outcome of a competitive equilibrium. An allocation is implementable if it satisfies Definition 1.

Definition 1. *An allocation $\{c_t, l_t, g_t, k_{t+1}\}_{t=0}^{\infty}$ is implementable with proportional taxes on capital and labour income if it satisfies the following conditions:*

- *There exists a sequence of taxes $\{\tau_{t+1}^K, \tau_t^L\}_{t=0}^{\infty}$, factor prices $\{w_t, r_{t+1}\}_{t=0}^{\infty}$ and asset holdings $\{a_{t+1}\}_{t=0}^{\infty}$ such that the allocation solves the individual's problem, given the prices;*
- *There exist factor prices $\{w_t, r_{t+1}\}_{t=0}^{\infty}$, such that the firm maximises its profits every period;*
- *The allocation satisfies the government budget constraint* (2.8) *every period;*
- *The allocation satisfies the aggregate resource constraint* (2.13) *every period;*
- *The allocation satisfies the domestic capital market equilibrium condition* (2.14) *every period.*

The next step is to derive the implementability constraint. First, use the individual's first order conditions (2.6) and (2.7) to substitute out the net prices in the individual's budget constraint (2.2). Multiply the result by $\beta^t u_{c_t}$, sum over the individual's lifetime and use the transversality condition

[12]Furthermore, the transversality condition for capital must hold: $\lim_{t\to\infty} \frac{k_{t+1}}{\prod_{s=1}^{t}(1+r_s)} = 0$.

for private assets (2.3) to find:

$$\sum_{t=0}^{\infty} \beta^t (c_t u_{c_t} + l_t u_{l_t}) = 0. \tag{2.15}$$

Note that there is no term in equation (2.15) which is associated with the period-0 term in Chamley (1986), since we assumed that no capital is required for period 0 production. Thus, the tax system is complete because the government can control all choice margins (i.e. all labour supply and saving decisions) with linear taxes on labour income and capital income. Therefore, capital income is not taxed (in the short run) to remedy an incompleteness in the tax code as in Chamley (1986).

Lemma 2.1 shows that an allocation that satisfies the implementability (2.15) and aggregate resource constraints (2.13) is implementable with proportional taxes on capital and labour income. Therefore, instead of directly choosing the optimal taxes (the dual problem), we can solve the government's problem by choosing the implementable allocation that maximises the representative individual's utility (the primal problem). We can then use the optimal allocation to retrieve the optimal tax rules.

Lemma 2.1. *An allocation is implementable with proportional taxes if and only if it satisfies the implementability constraint* (2.15) *and the aggregate resource constraint* (2.13).

Proof. See Appendix 2.A. □

2.2.6 Optimal taxation

In order to simplify notation, we denote the multiplier on the implementability constraint (2.15) by θ and define a pseudo utility function $W(\cdot)$ as:

$$W(c_t, l_t, \theta) \equiv u(c_t, l_t) + \theta(u_{c_t} c_t + u_{l_t} l_t). \tag{2.16}$$

$W(c_t, l_t, \theta)$ can be interpreted as the net social value of private utility, where the multiplier θ is a measure of aggregate tax distortions. We can then summarise the government problem as follows:

$$\max_{\{c_t, l_t, k_{t+1}\}_{t=0}^{\infty}} \quad \sum_{t=0}^{\infty} \beta^t W(c_t, l_t, \theta),$$

subject to

$$\begin{aligned} & c_0 + g_0 + k_1 = f_0(l_0), \\ & c_t + g_t + k_{t+1} - (1-\delta)k_t = f(k_t, l_t), \quad t > 0, \\ & \lim_{t\to\infty} \frac{k_{t+1}}{\prod_{s=1}^{t}(1+r_s)} = 0. \end{aligned} \tag{2.17}$$

We obtain the following first-order conditions for the government problem:

$$-\frac{W_{l_t}}{W_{c_t}} = f_{l_t} = w_t, \quad t \geq 0, \tag{2.18}$$

$$\frac{W_{c_t}}{\beta W_{c_{t+1}}} = 1 + f_{k_{t+1}} - \delta = 1 + r_{t+1}, \quad t \geq 0. \tag{2.19}$$

Equation (2.18) is the counterpart of the individual's first-order condition for labour supply (2.6). The government chooses the amount of labour in the economy until the social marginal utility cost of working $-W_{l_t}$ equals the social marginal benefit of working $w_t W_{c_t}$. Equation (2.19) is the government's Euler equation for consumption, which is the counterpart of the individual's Euler equation (2.7). The government chooses the consumption path such that the marginal decrease in social welfare incurred when saving in the current period W_{c_t} is equal to the marginal increase in social welfare from consuming the proceeds of the savings in the next period $(1 + r_{t+1})W_{c_{t+1}}$.

By taking derivatives of W in (2.16), we can find expressions for W_{c_t} and W_{l_t}:

$$W_{c_t} = u_{c_t}\left(1 + \theta + \theta\left(\frac{u_{c_t c_t} c_t}{u_{c_t}} + \frac{u_{c_t l_t} l_t}{u_{c_t}}\right)\right), \tag{2.20}$$

$$W_{l_t} = u_{l_t}\left(1 + \theta + \theta\left(\frac{u_{c_t l_t} c_t}{u_{l_t}} + \frac{u_{l_t l_t} l_t}{u_{l_t}}\right)\right). \tag{2.21}$$

We define the general-equilibrium elasticities ε_t^c and ε_t^l as:

$$-\frac{1}{\varepsilon_t^c} \equiv \frac{u_{c_t c_t} c_t}{u_{c_t}} + \frac{u_{c_t l_t} l_t}{u_{c_t}} = \frac{\partial \ln u_{c_t}}{\partial \ln c_t} + \frac{\partial \ln u_{c_t}}{\partial \ln l_t}, \tag{2.22}$$

$$-\frac{1}{\varepsilon_t^l} \equiv \frac{u_{c_t l_t} c_t}{u_{l_t}} + \frac{u_{l_t l_t} l_t}{u_{l_t}} = \frac{\partial \ln u_{l_t}}{\partial \ln c_t} + \frac{\partial \ln u_{l_t}}{\partial \ln l_t}. \tag{2.23}$$

The term ε_t^c captures the distortions in consumption and labour supply caused by changes in u_{c_t}, which in equilibrium equals the price of consumption. The capital income tax raises the price of consumption at date $t+1$ relative to consumption at date t. Hence, it induces substitution away from future consumption and future leisure towards current consumption and current leisure. Similarly, the term ε_t^l captures distortions in consumption and labour supply caused by changes in u_{l_t}, which is in equilibrium equal to the price of labour. The next proposition derives the optimal capital income tax in a given period t.

Proposition 2.1. *The optimal linear taxes on capital and labour income in each period are, respectively:*

$$\frac{r_{t+1}\tau_{t+1}^K}{1 + r_{t+1}} = \frac{\theta(1/\varepsilon_{t+1}^c - 1/\varepsilon_t^c)}{1 + \theta - \theta/\varepsilon_t^c}, \quad t > 0, \tag{2.24}$$

$$\frac{1}{1-\tau_t^L} = \frac{1+\theta-\theta/\varepsilon_t^l}{1+\theta-\theta/\varepsilon_t^c}, \quad t>0. \tag{2.25}$$

Proof. Substitute the expression for W_{c_t} from equation (2.20) into the government's Euler equation (2.19) and use the individual's Euler equation (2.7) to establish the optimal capital income tax. Similarly, substitute the expressions for W_{c_t} and W_{l_t} from equations (2.20) and (2.21) into (2.18) and use the first-order conditions for the household in (2.5) and (2.6) to find the optimal labour tax. □

Proposition 2.1 shows that taxes on capital income are desirable only if the aggregate elasticity today ε_t^c is higher than the aggregate elasticity tomorrow ε_{t+1}^c. Equivalently, capital income should be taxed only if the combined distortions in consumption demand and labour supply tomorrow are lower than the combined distortions in both consumption and labour today. Similarly, labour income taxes should be positive if the aggregate elasticity of present consumption is larger than that of present labour. This conforms to standard Ramsey intuitions. In Corollary 2.1, we show how it is optimal to set capital income taxes to zero in the steady-state.[13]

Corollary 2.1. *In the steady state, the optimal capital income tax is zero:* $\tau^K = 0$.

Proof. In the steady state, both consumption and leisure become constant, so ε_t^c becomes constant. From Proposition 1, it follows that $\tau_t^K = 0$ in the steady state. □

2.2.7 Why is the long-run tax on capital income zero?

We argue that in the steady state, the taxation of capital income should follow the prescriptions from the literature on optimal commodity taxation. In our model, a positive tax on capital income is equivalent to taxing future consumption at a higher rate than present consumption. Similarly, a zero capital income tax is equivalent to a uniform commodity tax on consumption at different dates. Corlett and Hague (1953) show that commodity tax differentiation is generally desirable because the distortions in commodity demands help alleviate distortions in labour supply. Conversely, if differentiated commodity taxes cannot mitigate labour supply distortions, they should be uniform, in order to avoid distortions in commodity demands.

We analyze a marginal tax reform to demonstrate why capital income taxes are only useful to alleviate labour market distortions and should be set to 0 in the steady state.[14] The policy experiment raises the capital income tax at time $t+1$ such that consumption at time t increases with ϵ and consumption at time $t+1$ declines with an amount to be yet determined. The policy experiment

[13] Moreover, this result ensures that the transversality condition for government debt holds ex-post. Since $r = (1-\tau^K)r$ when $\tau^K = 0$, the capital market equilibrium condition (2.14) holds and the transversality conditions for private assets and capital hold. Hence, the transversality condition for government debt will hold automatically.

[14] See also Albanesi and Armenter (2009) who employ a similar perturbation to argue that front-loading tax distortions is desirable.

keeps the entire intertemporal allocation at all dates $v \neq t, t+1$ unchanged. Hence, capital stocks at all dates t, except at date $t+1$, remain constant. Moreover, the policy experiment respects the implementability constraint. Therefore, taxes on labour income in period t and $t+1$ adjust to ensure that the intertemporal allocation remains constant and the implementability constraint is respected. Government spending does not change. We calculate the welfare effects of this small tax perturbation and show that they are critically determined by the responses of labour supply to the capital income tax. Since the allocation for all periods except t and $t+1$ does not change, raising the capital income tax in period $t+1$ only affects utility $\mathcal{W}$ in periods t and $t+1$:

$$\mathcal{W} \equiv u(c_t, l_t) + \beta u(c_{t+1}, l_{t+1}), \quad t > 0. \tag{2.26}$$

The next proposition derives the welfare effects of this tax perturbation.

Proposition 2.2. *Starting from a given initial allocation, the welfare effect of marginally raising the capital income tax such that c_t increases with ϵ, while respecting the resource and implementability constraints by adjusting the taxes on labour income, is given by:*

$$\begin{aligned}\frac{\mathrm{d}\mathcal{W}}{u_{c_t}} &= -(1-\tau_t^L)w_t\left(1-\frac{1}{\alpha_t}\right)\mathrm{d}l_t - \frac{(1-\tau_{t+1}^L)w_{t+1}}{(1+(1-\tau_{t+1}^K)r_{t+1})}\left(1-\frac{1}{\alpha_{t+1}}\right)\mathrm{d}l_{t+1} \\ &= (1-\alpha_t)\,\epsilon - \left(\frac{1+r_{t+1}}{1+(1-\tau_{t+1}^K)r_{t+1}}\right)\left(\frac{1-\frac{\alpha_t}{1-\tau_t^L}}{1-\frac{\alpha_{t+1}}{1-\tau_{t+1}^L}}\right)(1-\alpha_{t+1})\,\epsilon,\end{aligned} \tag{2.27}$$

$$\alpha_t \equiv \frac{1+\frac{c_t u_{c_t c_t}}{u_{c_t}}+\frac{l_t u_{l_t c_t}}{u_{c_t}}}{1+\frac{c_t u_{c_t l_t}}{u_{l_t}}+\frac{l_t u_{l_t l_t}}{u_{l_t}}} = \frac{1-\frac{1}{\varepsilon_t^c}}{1-\frac{1}{\varepsilon_t^l}}, \quad t > 0. \tag{2.28}$$

Proof. The tax reform should keep the intertemporal allocation of resources constant and must be implementable with linear taxes on capital income in period $t+1$ and linear taxes on labour income in periods t and $t+1$. First, this requires that the reform respects both the resource constraints in periods t and $t+1$:

$$f\left(\bar{k}_t, l_t\right) = c_t + \bar{g}_t + k_{t+1} - (1-\delta)\,\bar{k}_t, \tag{2.29}$$

$$f\left(k_{t+1}, l_{t+1}\right) = c_{t+1} + \bar{g}_{t+1} + \bar{k}_{t+2} - (1-\delta)\,k_{t+1}, \tag{2.30}$$

where a bar indicates a variable that does not change under the reform. Second, the tax reform should respect the implementability constraints in periods t and $t+1$:

$$c_t u_{c_t} + l_t u_{l_t} + \beta(c_{t+1}u_{c_{t+1}} + l_{t+1}u_{l_{t+1}}) = \zeta_t, \tag{2.31}$$

for some exogenous value ζ_t of the implementability constraints in all periods t. Since we can adjust taxes on labour income in both period t and period $t+1$, we can construct a policy reform such that the change in the implementability constraints in both period t and period $t+1$ is zero by appropriate changes in the taxes on labour income:

$$\mathrm{d}(u_{c_t} c_t + u_{l_t} l_t) = 0, \tag{2.32}$$

$$\mathrm{d}(u_{c_{t+1}} c_{t+1} + u_{l_{t+1}} l_{t+1}) = 0. \tag{2.33}$$

Note that if the policy experiment satisfies (2.32) and (2.33), then the implementability constraint (2.31) is respected.

The policy experiment raises consumption c_t at time t with $\mathrm{d}c_t = \epsilon$. The change in labour supply l_t at time t follows from totally differentiating the period t implementability constraint (2.32):

$$\mathrm{d}l_t = -\alpha_t \frac{u_{c_t}}{u_{l_t}} \mathrm{d}c_t = -\alpha_t \frac{u_{c_t}}{u_{l_t}} \epsilon, \tag{2.34}$$

where α_t is defined in Proposition 2.2. By noting that k_t is predetermined at time t, the change in k_{t+1} is found by totally differentiating the period t resource constraint (2.29):

$$\mathrm{d}k_{t+1} = f_l \mathrm{d}l_t - \mathrm{d}c_t = -\left(1 + f_l \alpha_t \frac{u_{c_t}}{u_{l_t}}\right) \epsilon, \tag{2.35}$$

where the last part follows upon substitution of $\mathrm{d}l_t = -\alpha_t \frac{u_{c_t}}{u_{l_t}} \epsilon$ and $\mathrm{d}c_t = \epsilon$. Similarly, the policy reform lowers consumption c_{t+1} at time $t+1$. By totally differentiating the period $t+1$ implementability constraint (2.33), we find the change in labour $\mathrm{d}l_{t+1}$ at time $t+1$:

$$\mathrm{d}l_{t+1} = -\frac{u_{c_{t+1}}}{u_{l_{t+1}}} \alpha_{t+1} \mathrm{d}c_{t+1}. \tag{2.36}$$

By differentiating the economy's resource constraint at $t+1$ in (2.30), we find the change in consumption $\mathrm{d}c_{t+1}$ at $t+1$ (note that k_{t+2} does not change):

$$\mathrm{d}c_{t+1} = (f_{k_{t+1}} + 1 - \delta)\mathrm{d}k_{t+1} + f_{l_{t+1}} \mathrm{d}l_{t+1} = -(f_{k_{t+1}} + 1 - \delta) \left(\frac{1 + f_{l_t} \frac{u_{c_t}}{u_{l_t}} \alpha_t}{1 + f_{l_{t+1}} \frac{u_{c_{t+1}}}{u_{l_{t+1}}} \alpha_{t+1}} \right) \epsilon, \tag{2.37}$$

where the second part follows upon substitution of equations (2.35) and (2.36). Consequently, we find for $\mathrm{d}l_{t+1}$:

$$\mathrm{d}l_{t+1} = \frac{u_{c_{t+1}}}{u_{l_{t+1}}} \alpha_{t+1} (f_{k_{t+1}} + 1 - \delta) \left(\frac{1 + f_{l_t} \frac{u_{c_t}}{u_{l_t}} \alpha_t}{1 + f_{l_{t+1}} \frac{u_{c_{t+1}}}{u_{l_{t+1}}} \alpha_{t+1}} \right) \epsilon. \tag{2.38}$$

Totally differentiating (2.26) gives the change in social welfare:

$$\mathrm{d}\mathcal{W} = u_{c_t}\mathrm{d}c_t + u_{l_t}\mathrm{d}l_t + \beta u_{c_{t+1}}\mathrm{d}c_{t+1} + \beta u_{l_{t+1}}\mathrm{d}l_{t+1}. \tag{2.39}$$

Substitute for the changes consumption using equations (2.34) and (2.36) to find:

$$\frac{\mathrm{d}\mathcal{W}}{u_{c_t}} = \frac{u_{l_t}}{u_{c_t}}\left(1 - \frac{1}{\alpha_t}\right)\mathrm{d}l_t + \frac{\beta u_{c_{t+1}}}{u_{c_t}}\frac{u_{l_{t+1}}}{u_{c_{t+1}}}\left(1 - \frac{1}{\alpha_{t+1}}\right)\mathrm{d}l_{t+1}. \tag{2.40}$$

Substituting the first-order conditions of the household in equations (2.5), (2.6) and (2.7) gives the first part of the proposition. Finally, we can substitute the changes in labour supply (2.34) and (2.36) into (2.40) and use the firm's first-order conditions in (2.10) and (2.11) to find the second part of the proposition. □

Consequently, Proposition 2.2 recovers the Corlett-Hague motive for capital taxation in the Chamley-Judd framework. The first part of the Proposition shows how an increase in the capital income tax lowers labour supply at t ($\mathrm{d}l_t < 0$) and increases labour supply at $t+1$ ($\mathrm{d}l_{t+1} > 0$). The reason for this rotation of the labour supply schedule over time is twofold, see also Jacobs and Schindler (2012). On the one hand, future leisure becomes more expensive in terms of current leisure, which leads to intertemporal substitution in leisure: labour supply in period $t+1$ increases and labour supply in period t decreases. These effects are associated with the ε_t^l-terms. On the other hand, capital income taxes also make future consumption more expensive relative to current consumption. The corresponding substitution effect implies that consumption in period $t+1$ decreases and consumption in period t increases. This latter change in consumption causes income effects in labour supply: lower consumption in period $t+1$ implies that labour supply in period $t+1$ increases, while higher consumption in period t implies that labour supply in period t decreases. These effects are associated with the ε_t^c-terms. If the increase (decrease) in labour supply at time $t+1$ (t) is sufficiently large (small), social welfare increases ($\mathrm{d}\mathcal{W} > 0$). Consequently, the increase in the capital income tax is socially desirable.

The α_t-terms ($\alpha_t = (1 - 1/\varepsilon_t^c)(1 - 1/\varepsilon_t^l)^{-1}$) capture the complementarity between consumption and labour. If $\alpha_{t+1} < \alpha_t$, consumption at date $t+1$ is less complementary with labour in periods t and $t+1$ than consumption at date t. Consequently, introducing a capital income tax is socially desirable, provided there is no initial capital income taxation (i.e. $\tau_{t+1}^K = 0$) and labour taxes are constant over time (i.e. $\tau_t^L = \tau_{t+1}^L$). If there is a positive pre-existing capital income tax ($\tau_{t+1}^K > 0$), increasing it further is socially desirable only if the benefits of reduced labour market distortions are still larger than the costs of larger saving distortions. Clearly, if labour taxes are not constant over time, intertemporal labour supply decisions are distorted. Then, the capital income tax can either alleviate or exacerbate the intertemporal labour market distortions generated by non-constant labour taxes. The latter finding has not yet received a lot of attention in the literature: zero optimal capital taxation generally requires constant labour taxes. If, for whatever reason, labour taxes are not constant, optimal capital income taxes need not be zero.

To further illustrate the Corlett-Hague motive, we can analyze the welfare effect of introducing a small capital income tax in a setting with constant labour taxation ($\tau_t^L = \tau_{t+1}^L = \tau^L$) and no initial capital income taxation (τ_{t+1}^K). Using equation (2.27), it follows that the welfare effect of such a reform is:

$$\frac{\mathrm{d}\mathcal{W}}{u_{c_t}} = \frac{(\alpha_t - \alpha_{t+1})\,\tau^L}{1 - \tau^L - \alpha_{t+1}}\epsilon, \quad t > 0. \tag{2.41}$$

Thus, the introduction of a capital income tax is socially desirable only if consumption at date $t+1$ is less complementary to labour than consumption at date t, i.e. if $\alpha_{t+1} < \alpha_t$.

Proposition 2.2 governs the desirability of capital income taxation even if the economy has not converged to a steady state. In the steady state, c, l and k are all constant, which renders α, f_k, f_l, u_c and u_l constant. If labour taxes are constant, we can use equation (2.27) to calculate the welfare effect of raising the capital income tax in the steady state:

$$\frac{\mathrm{d}\mathcal{W}}{u_c} = -\frac{(1-\alpha)\,\tau_K r}{1 + (1-\tau^K)r}\epsilon < 0. \tag{2.42}$$

Raising the capital income tax in a steady state with constant labour taxes unambiguously lowers social welfare: the increase in distortions from lower current labour supply is larger than the decrease in distortions in future labour supply. Only if the initial capital tax is zero, i.e. $\tau_K = 0$, then the welfare effect of raising the capital tax is zero, i.e. $\mathrm{d}\mathcal{W} = 0$. Hence, the optimal tax on capital income in the steady state is zero. The implication is clear: capital income taxes are not desirable.

The following Corollary demonstrates that the optimal capital income tax derived under the perturbation approach is exactly the same as the optimal capital income tax derived in Proposition 2.1, provided labour income taxes are optimised. Thus, the perturbation approach leads to the same solution as the primal approach.

Corollary 2.2. *If labour taxes are optimized according to* (2.25)*, the perturbation approach gives the following optimal capital income tax:*

$$\frac{r_{t+1}\tau_{t+1}^K}{1 + r_{t+1}} = \frac{\theta(1/\varepsilon_{t+1}^c - 1/\varepsilon_t^c)}{1 + \theta - \theta/\varepsilon_t^c}, \quad t > 0. \tag{2.43}$$

Proof. In the optimum, the marginal benefits and the marginal costs of the reform should cancel out, so $\mathrm{d}\mathcal{W} = 0$ in equation (2.27). Rewriting that expression, we obtain:

$$\frac{1+(1-\tau_{t+1}^K)r_{t+1}}{1+r_{t+1}} = \frac{1-\frac{\alpha_t}{1-\tau_t^L}}{1-\frac{\alpha_{t+1}}{1-\tau_{t+1}^L}} \frac{1-\alpha_{t+1}}{1-\alpha_t} \tag{2.44}$$

Substituting α_t and α_{t+1} from equation (2.28) and τ_t^L and τ_{t+1}^L from the optimal labour income tax expression in equation (2.25), we obtain the optimal capital income tax τ_{t+1}^K:

$$\frac{r_{t+1}\tau_{t+1}^K}{1+r_{t+1}} = \frac{\theta(1/\varepsilon_{t+1}^c - 1/\varepsilon_t^c)}{1+\theta-\theta/\varepsilon_t^c} \tag{2.45}$$

The expression above is identical to the expression obtained using the primal approach in equation (2.24). □

2.3 Interpretations in the literature

The optimal taxation literature discusses two main economic intuitions that would explain the Chamley-Judd result that the tax on capital income should be zero in the long run. The first is that a non-zero capital income tax results in exploding tax distortions in finite time, which violates the Ramsey principle to smooth distortions over time, see also Judd (1999) and Banks and Diamond (2010). The second intuition is that if the supply of capital is infinitely elastic in the long run, all taxes are borne by labour in any case. Hence, it is better not to distort capital accumulation by setting a zero tax on capital income, see also Auerbach and Kotlikoff (1983), Correia (1996) and Mankiw, Weinzierl, and Yagan (2011). This section argues that the first intuition can be interpreted as a special case of our generalized Corlett-Hague intuition and the second intuition is misleading.

2.3.1 Intuition 1: exploding tax distortions

Can the Chamley-Judd results be interpreted as a strict application of the Ramsey principle, as in Judd (1999) and Banks and Diamond (2010)? In this section, we show how the Ramsey intuition of taxing inelastic consumption demands at higher rates can be seen as a special case of the Corlett-Hague intuition which calls for taxing leisure complements at higher rates (Corlett and Hague, 1953). Moreover, the Chamley-Judd result can be seen as an application of the Ramsey principle only when restrictive assumptions are made on the utility function.

We can gain more intuition as to how the standard mechanisms from the static models in the optimal taxation literature apply to the dynamic model developed in this paper. We can measure the complementarity between consumption at period t and leisure at period j by $\varepsilon^*_{c_t w^*_j}$, which is the compensated elasticity of consumption c_t with respect to the net wage $w^*_j \equiv (1-\tau^L_j)w_j$ in period j, see also Diamond and Mirrlees (1971b), Sandmo (1974) and Atkinson and Stiglitz (1980). A compensated increase in the net wage w^*_j leads to an increase in labour l_j, or alternatively, a decrease in leisure $1-l_j$. If $\varepsilon^*_{c_{t+1} w^*_j} > \varepsilon^*_{c_t w^*_j}$, then the increase in w_j leads to a larger increase in c_{t+1} than in c_t. This implies that c_{t+1} is more complementary to labour than c_t or, equivalently, c_{t+1} is less complementary to leisure than c_t in the Corlett-Hague sense. Similarly, we define the compensated price elasticity of consumption with respect to the net interest rate $r^*_j \equiv (1-\tau^K_j)r_j$ in period j as $\varepsilon^*_{c_t r_j}$. In Proposition 2.3, we show that if consumption demands depend solely on contemporaneous prices, i.e. the net interest rate and wage rate in that period, the goods that are less price elastic are also the consumption goods that are relatively more complementary to leisure.

Proposition 2.3. *Assume that there exists a final time period T. If consumption demands depend only on prices in period t, and consumption in period t is more elastic with respect to the net interest rate than consumption in period $t+1$, so that $\varepsilon^*_{c_t r^*_t} > \varepsilon^*_{c_{t+1} r^*_{t+1}}$, then consumption in period t is also more complementary to leisure than consumption in period $t+1$, i.e. $\varepsilon^*_{c_t w^*_t} < \varepsilon^*_{c_{t+1} w^*_{t+1}}$, since $\varepsilon^*_{c_t r^*_t} + \varepsilon^*_{c_t w^*_t} = 0$.*

Proof. Assume that there exists a final time period, T. This allows us to inspect the individual's expenditure minimisation problem, where the individual chooses consumption and leisure to minimise the lifetime income that attains utility $\bar{U}$. The individual's dual problem becomes:

$$\min_{\{c_t, l_t\}_{t=0}^T} c_0 + w^*_0(1-l_0) + \sum_{t=1}^{T} \frac{c_t + w^*_t(1-l_t)}{\prod_{s=1}^{t}(1+r^*_s)}, \tag{2.46}$$

$$\text{subject to } U(c_0,\dots,c_T,l_0,\dots,l_T) \geq \bar{U}. \tag{2.47}$$

Solving the problem above leads to compensated demands $\{c^*_t, l^*_t\}_{t=0}^T$, which are homogeneous of degree 0:

$$y_t(r^*_1,\dots,r^*_T,w^*_0,\dots,w^*_T) = y_t(\phi r^*_1,\dots,\phi r^*_T,\phi w^*_0,\dots,\phi w^*_T), \quad \phi > 0, \quad y_t = \{c_t, l_t\}. \tag{2.48}$$

We can differentiate this equation with respect to ϕ, set ϕ to 1, and define $\varepsilon^*_{y_t p_j}$ as the compensated elasticity of period t good $y_t = \{c_t, l_t\}$ with respect to perid j price $p_j = \{r^*_j, w^*_j\}$. This leads to:

$$\sum_{j=1}^{T} \varepsilon^*_{c_t r^*_j} + \sum_{j=0}^{T} \varepsilon^*_{c_t w^*_j} = 0 \tag{2.49}$$

If we assume that consumption elasticities solely depend on prices in period t, this expression collapses to the proposition. □

Proposition 2.3 shows that if consumption demands solely depend on contemporaneous prices, a good that is very elastic with respect to its own price will also be very complementary to leisure: thus high elasticities of consumption with respect to net interest rates $\varepsilon^*_{c_t r^*_j}$ mean low compensated elasticities of consumption with respect to net wage rates $-\varepsilon^*_{c_t w^*_t}$ and vice versa. Consequently, the Ramsey inverse-elasticity rule is nested as a special case of the general Corlett-Hague rule for commodity taxation. This can also be seen from the definition of the general-equilibrium elasticity ε^c_t in equation (2.22). Naturally, if the utility function is separable, so that $u_{cl} = 0$ in equation (2.22), the Ramsey intuition is applicable. In this pure Ramsey case, capital income is taxed only if the elasticity of marginal utility of consumption $\left(\frac{\partial \ln c_t}{\partial \ln u_{c_t}}\right)^{-1}$ varies with time.

However, the standard Ramsey intuition – that inelastic goods should be taxed at higher rates – need not always be applicable: the Ramsey explanation critically depends on the assumption that compensated demands for goods depend solely on contemporaneous prices. If compensated demands also depend on other prices, it is theoretically possible to have a good that is both inelastic with respect to its own price and is complementary to labour at the same time. In that case, it could be that the complementarity is so strong that it becomes optimal to subsidise the good to reduce labour supply distortions. To see why, in the general case the general-equilibrium elasticity ε^c_t includes complementarities with labour, i.e. $\frac{\partial \ln l_t}{\partial \ln u_{c_t}}$ that are not present in the own-price elasticities, i.e. $\frac{\partial \ln c_t}{\partial \ln u_{c_t}}$. By distorting the consumption prices, the capital income tax not only distorts the intertemporal allocation of consumption, but also affects the intertemporal allocation of labour supply. Given that labour supply is distorted by the labour income tax, a capital income tax (or a subsidy) can be helpful to reduce labour supply distortions. This depends on the specific pattern of $\frac{\partial \ln l_t}{\partial \ln u_{c_t}}$ over time and no general conclusion can be drawn about this term without imposing further structure on the utility function.

In order to prove Proposition 2.3, we assumed the existence a final period T. This is a technical assumption that ensures we can analyse the individual's dual problem without focusing on the issue of infinite commodity spaces. The result in Proposition 2.3 is valid for an arbitrarily large T, so the assumption of finite time should not obscure the relevance of the Proposition.

2.3.2 No convergence to steady state needed?

Our analysis so far suggests that capital income taxes are optimally zero in a limited array of cases, namely if the economy is in a steady-state, or if preferences are restricted to a specific class of utility functions. However, Judd (1999) argues that under any utility function, distortions arising from capital income taxation would explode in finite time. Thus, the optimal tax on capital income would be driven down to zero as the deadweight loss of taxation would reach an upper bound in finite time. Hence, the optimal tax on capital income is zero in finite time even if the economy does not converge to a steady-state. This finding seems to suggest that no restrictions on the utility function are needed to obtain a zero tax on capital income in finite time.

However, Judd (1999) does not take into account that taxes on capital income may be desirable to alleviate the distortions of taxes on labour income on labour supply. While the wedge between the MRS and MRT between consumption at early periods and future consumption can indeed grow at an exponential rate if capital income is taxed, this can be optimal if labour supply distortions would also grow exponentially over time. Hence, one cannot a priori conclude that capital income taxes should converge to zero in finite time[15]. From the models with a finite time horizon, we know that the Corlett-Hague motive is generally present unless restrictions are imposed on the utility function, see for example Atkinson and Sandmo (1980) and Erosa and Gervais (2002).

Moreover, the analysis of Judd (1999) also reveals that the deadweight loss of taxation becomes constant in finite time only if the general-equilibrium elasticity ε_t^c converges to a constant in finite time, see his equation (28). He then concludes that the steady-state is not required to obtain a zero capital income tax: the result holds in finite time, as long as the bound on the multipliers holds. Straub and Werning (2020) show that this result needs a large qualification, as Judd (1999) assumes that the endogenous multipliers of the government's budget constraint are bounded. We agree with the qualifications raised by Straub and Werning (2020). However, we take their argument further: we look at the case where this bound is not required, namely when preferences are such that the analysis of Judd (1999) is valid. In particular, the assumptions needed to ensure that distortions reach an upper bound in Judd (1999) are equivalent to assuming that utility is time-separable, separable between consumption and labour and homothetic in consumption. In Corollary 2.3, we show that if preferences satisfy these assumptions, ε_t^c is constant in all periods, not just in the steady state.

Corollary 2.3. *If the utility function is additively time-separable, strongly separable between consumption and labour and homothetic in the consumption sub-utility, then ε_t^c is constant and capital income taxes are optimally zero at all dates.*

Proof. See Chari and Kehoe (1999) and Appendix 2.C. □

Intuitively, these assumptions on preferences ensure that the Corlett-Hague motive for taxing capital income vanishes. Due to the preference structure assumed in Corollary 2.3, the general equilibrium

[15] A similar argument is put forward in Straub and Werning (2020), who show that the ratio of the marginal costs and benefits of taxation remains constant. This suggests that discussing only the costs of taxation does not give a complete picture of the trade-offs faced by the government.

elasticity ε_t^c is constant, so consumption is equally complementary to leisure in every period. This makes capital income taxes redundant in every period and not only in the steady-state. This becomes immediately apparent in equation (2.22): if the utility function is separable between consumption and labour, $u_{cl} = 0$ and the second term of the equation is zero. Furthermore, if the consumption sub-utility is homothetic, the first term of equation (2.22) becomes a constant. The combination of these two properties renders ε_t^c constant. Thus, the argument in Judd (1999) that no steady-state is needed for capital income taxation to be zero is equivalent to our argument that capital income taxes are zero because the Corlet-Hague complementarity motive vanishes. Assuming preferences are such that distortions reach an upper bound is equivalent to assuming time separability, separability between consumption and leisure and homotheticity of the consumption sub-utility.

Corollary (12) of Judd (1999) demonstrates that if the assumptions of separability between consumption and leisure and homotheticity of the consumption sub-utility are violated, the optimal tax on capital income is not zero if the steady-state is not reached. In particular, assuming a Stone-Geary utility function that is separable between consumption and labour, Judd (1999) concludes that "the capital income tax is never zero, but for reasons which are consistent with the inverse-elasticity rule", i.e. the Ramsey rule. In this case, the term $\frac{\partial \ln c_t}{\partial \ln u_{c_t}}$ in equation (2.22) is never constant and u_{cl} equals 0. Thus, without invoking the steady-state assumption, or without assuming separable and homothetic preferences, the capital income tax rate fluctuates according to the inverse of the elasticity of consumption, i.e. according to whether consumption is more or less complementary to leisure over time.

To conclude, the standard Ramsey intuition applied in Judd (1999) and Banks and Diamond (2010) need not always be applicable: this critically depends on the general-equilibrium elasticity ε_t^c converging to a constant, which either requires specific assumptions on the utility function (namely, separability between consumption and leisure and homotheticity of the consumption sub-utility) or convergence to a steady state.

2.3.3 Non-separable utility

So far, we focused solely on optimal capital income taxation if utility is additively separable with respect to time. A natural question then arises: how should capital income be taxed if utility is not time-separable? Straub and Werning (2020) showed the importance of the assumption of time separability in the analysis of Chamley (1986). If the individual utility function is not time-separable, convergence to a steady-state is unlikely and, even if it occurs, the steady-state features either zero private wealth or a first-best outcome. In such cases, it can be optimal to indefinitely tax capital income at the maximum rate. The results obtained by Straub and Werning (2020) with non-additive utility are mostly of theoretical interest, since first-best outcomes are unlikely to occur in practice and 100% capital income taxes are not implementable in market economies.

Our main intuition nevertheless carries over to the non-separable case. The next Lemma shows the conditions necessary for the capital income taxes to be zero if preferences are not time-separable and the individual faces a finite horizon. No steady-state assumptions are invoked here.

Lemma 2.2. *If the agent faces a finite horizon $0 < t \leq T$, and preferences are of the form:*

$$U = U(h(c_0, \ldots, c_T), v(l_0, \ldots, l_T)),$$

with $h(\cdot)$ and $v(\cdot)$ denoting homothetic sub-utility functions, then the capital income taxes are optimally zero in every period.

Proof. See Appendix 2.B. □

If utility is weakly separable between consumption and leisure and homothetic both in consumption and leisure, there is no scope for capital income taxes. The intuition for the result is the same as in the time-separable case: the weak separability and homotheticity of the utility function makes present and future consumption equally complementary to leisure, rendering capital income taxes ineffective for alleviating labour supply distortions.

2.3.4 Intuition 2: full tax shifting to labour

Another common explanation for the zero optimal capital income tax result can be found in the work of Auerbach and Kotlikoff (1983), Correia (1996) and Mankiw, Weinzierl, and Yagan (2011). These authors assert that the supply of capital becomes infinitely elastic in the long run, so that the entire burden of a tax on capital income is borne by labour through factor price adjustments. Since in the long run the net interest rate is fixed by exogenous factors – such as the rate of time preference and depreciation – any decrease due to taxing capital income will be perfectly offset by a one-to-one increase in the gross interest rate. To achieve the decrease in the gross interest rate, capital stock must decrease, which leads to a decrease in gross wages.

While we agree that the infinite elasticity of capital supply is a feature of our standard neoclassical model, we believe that factor price adjustments cannot be the driving force behind the Chamley-Judd result. To show this, we switch off the general-equilibrium effects on the interest rate by considering the case of an open economy in this section. Since the gross interest rate is fixed in the world asset markets, the tax burden on capital cannot be shifted towards labour through general-equilibrium effects on factor prices. If the reason capital income taxes are zero is that all tax burden is shifted to labour due to general-equilibrium effects in factor prices, capital income taxes should not be zero when there are no such general-equilibrium effects. However, we show that capital income taxes remain zero in the steady-state, despite the absence of factor price adjustments.

The representative individual and the government are allowed to access a perfectly competitive international capital market in which a foreign asset x_t is traded. Foreign capital x_t is supplied infinitely elastically and yields an exogenously given return r_t, which is the required return for private debt a_t and government debt d_t. The government and the representative individual have optimisation problems that are identical to closed-economy case, with the only difference that now both have access to the international capital market. Moreover, the implementability constraint remains identical to the one derived in (2.15).

The main difference with the closed-economy set-up is the assumption that the production technology $f(\cdot)$ employs only labour: $f(l_t) = A_t l_t$. This way, we sever the link between wages and interest rates, while keeping everything else identical with the closed-economy setting. Profit maximisation then implies that labour demand is perfectly elastic at the market wage: $A_t = w_t$.

In this open economy, total domestic production $w_t l_t$ need not equal domestic absorption $c_t + g_t$. Hence, the current account is determined by:

$$c_t + g_t + x_{t+1} - (1 + r_t)x_t \leq w_t l_t, \quad t \geq 0, \tag{2.50}$$

$$\lim_{t\to\infty} \frac{x_{t+1}}{\prod_{s=1}^{t}(1 + r_s)} = 0. \tag{2.51}$$

To prevent explosive paths of net foreign debt, we impose a no-Ponzi-game condition: current account deficits are always repaid with later current account surpluses.

The capital market equilibrium condition in the open economy is similar to the one in the closed-economy model in Section 2.2:

$$a_t - d_t = x_t, \quad t > 0. \tag{2.52}$$

The left-hand side represents the demand for foreign capital: both the individual and the government demand assets in their intertemporal trades. The right-side of the equation represents the total supply of capital: foreign capital flows into the economy, which increases the intertemporal consumption possibilities compared to the case when the economy is closed.

Since the domestic firm does not employ capital in its production process, we need to modify the definition of an implementable allocation to include the flows of foreign capital x_t instead of k_t, see Definition 2.

Definition 2. *An allocation $\{c_t, g_t, x_{t+1}\}_{t=0}^{\infty}$ is implementable with proportional taxes on capital and labour income, given the factor prices $\{w_t, r_t\}_{t=0}^{\infty}$, if it satisfies the following conditions:*

- *There exists the sequence of taxes $\{\tau_{t+1}^K, \tau_t^L\}_{t=0}^{\infty}$ and a sequence of asset holdings $\{a_{t+1}\}_{t=0}^{\infty}$ such that the allocation solves the individual's problem, given the prices;*
- *The allocation satisfies the government budget constraint* (2.8) *every period;*
- *The allocation satisfies the aggregate resource constraint* (2.50) *every period;*
- *The allocation satisfies the international capital market equilibrium* (2.52) *every period.*

Lemma 2.3 is the counterpart of Lemma 2.1 in an open-economy setting. It shows that for an allocation to be implementable with proportional taxes in an open-economy setting, it need only be feasible (satisfy the aggregate resource constraint (2.50)) and satisfy the implementability constraint (2.15).

Lemma 2.3. *An allocation $\{c_t, g_t, x_{t+1}\}_{t=0}^{\infty}$ is implementable with proportional taxes on capital and labour income, given the factor prices $\{w_t, r_{t+1}\}_{t=0}^{\infty}$, if and only if it satisfies the implementability constraint* (2.15) *and the aggregate resource constraint* (2.50).

Proof. See Appendix 2.D. □

Thus, if we denote by θ the multiplier on the implementability constraint (2.15) and define the modified welfare function as we did in the closed-economy case in equation (2.16), the optimisation problem of the government becomes:

$$\begin{aligned} \max_{\{c_t, l_t, x_{t+1}\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t W(c_t, l_t, \theta) \\ \text{subject to } & w_t, r_t \text{ given}, \quad \forall t, \\ & c_t + g_t + x_{t+1} - (1 + r_t)x_t = w_t l_t, \quad \forall t, \\ & \lim_{t\to\infty} \frac{x_{t+1}}{\prod_{s=1}^{t}(1 + r_s)} = 0. \end{aligned} \tag{2.53}$$

From this formulation it is obvious that the optimal tax problem in the open-economy setting is mathematically identical to the closed-economy setting. The only difference is cosmetic: the government chooses the amount of private domestic capital k_{t+1} in the closed economy and the amount of foreign capital x_{t+1} in the open economy. We thus confirm Diamond and Mirrlees (1971a): optimal tax expressions are identical the open economy with constant factor prices and in the closed economy with endogenous factor prices.[16] As a result, the steady-state optimal capital income tax expression will still lead to the Chamley (1986) and Judd (1985) result: $\tau^K = 0$.[17] Note that in the open economy, the real interest rate r is constant. Hence, a steady state only exists if an assumption is made on the discount factor β. In particular, the discount factor β must be must be consistent with the individual Euler condition in the steady-state (2.7): $\beta = [1 + (1 - \tau^K)r]^{-1}$.

One may wonder, then, to what extent the assumption on the discount rate assumes the zero tax on capital income? This is not the case. The government's Euler equation (2.19), together with the private Euler equation (2.7), simultaneously determine β and τ^K. This means that the open-economy assumption, i.e. fixing the gross interest rate, does not assume the zero tax result. The private Euler equation pins down a value for the discount rate β that is consistent with a steady state for any net interest rate (i.e. for any capital income tax τ^K, including a zero capital income tax). One then needs the government Euler equation to prove that the optimal capital income tax is indeed zero in the steady state.[18]

We have demonstrated that the optimal long-run capital income tax is zero both in open-economy and in closed-economy settings. This finding is not consistent with the notion that in the long run, the capital income tax is completely shifted to labour via general-equilibrium effects on interest rates and wages, as the latter are absent – by definition – in our open-economy setting.

2.4 Conclusion

This paper tried to answer the question: why is the long-run capital income tax zero in Chamley (1986) and Judd (1999)? We demonstrated that standard principles from the optimal commodity tax literature drive the result that capital income taxes should be zero. In particular, the steady-state assumption forces consumption at different dates to become equally complementary with leisure. This means that capital income taxes cannot be used to offset the labour supply distortions caused by labour income taxes.

Our interpretation of the results in Chamley (1986) and Judd (1999) is also consistent with subsequent results in the literature, which showed that the zero-tax result holds outside the steady-state, provided certain assumptions on preferences hold. We showed that these assumptions are equiva-

[16] We should note that the production efficiency theorem relies on the absence of pure profits (or the availability of a perfect profit tax) and full verifiability of all transactions between firms and households (Diamond and Mirrlees, 1971a). Our model satisfies both requirements.

[17] This result depends on the absence of pure profits (constant returns to scale in production) or the presence of a pure profit tax when returns to scale are not constant. See also Correia (1996).

[18] Similarly, in the closed economy the discount rate β is given and the interest rate r is endogenous. The private and government Euler equations then jointly determine the steady state gross interest rate r and the optimal capital income tax τ^K.

lent to assuming that the Corlett-Hague motive vanishes, as consumption and leisure become equally complementary throughout time. In doing so, we showed that the argument that a positive capital income tax would lead to exploding tax distortions in finite time needs reconsideration, as this intuition is applicable only when restrictions are made on the utility function. Furthermore, we showed that general-equilibrium effects on interest rates cannot be the main driver behind the long-run zero optimal capital income tax, since we found that the optimal capital income tax is zero also in an open-economy setting, where interest rates are constant.

2.A Proof of Lemma 2.1

Proof. We first prove that an implementable allocation satisfies the implementability constraint (2.15) and the aggregate resource constraint (2.13). Since an implementable allocation solves the individual's problem by definition, we can use the individual's optimality conditions, transversality condition for assets and the budget constraint to derive the implementability constraint. Furthermore, an implementable allocation satisfies the aggregate resource constraint by construction. Next, we prove that an allocation $\{c_t, l_t, g_t, k_{t+1}\}_{t=0}^{\infty}$ that satisfies (2.15) and (2.13) is implementable. We can start by defining the factor prices r_t and w_t such that the firm's optimality conditions hold:

$$r_t \equiv f_k(k_t, l_t) - \delta, \quad t > 0, \tag{2.54}$$

$$w_t \equiv f_l(k_t, l_t), \quad t \geq 0. \tag{2.55}$$

Given the factor prices, we can use the individual's first-order conditions to define the proportional taxes $\{\tau_t^L, \tau_{t+1}^K\}_{t=0}^{\infty}$ that implement the allocation $\{c_t, l_t, g_t, k_{t+1}\}_{t=0}^{\infty}$:

$$\tau_t^L \equiv 1 + \frac{u_{l_t}}{w_t u_{c_t}},, \quad t > 0, \tag{2.56}$$

$$\tau_{t+1}^K r_{t+1} \equiv 1 + r_{t+1} - \frac{u_{c_t}}{\beta u_{c_{t+1}}}, \quad t > 0. \tag{2.57}$$

Since we know the initial asset endowment $a_0 = 0$ and the paths of consumption and labour and the net prices of labour and future consumption, we can recursively define the private asset holdings $\{a_{t+1}\}_{t=0}^{\infty}$ such that the individual's budget constraint (2.2) holds:

$$a_{t+1} \equiv (1 + (1 - \tau_t^K)r_t) + (1 - \tau_t^L)w_t l_t - c_t, \quad t > 0, \tag{2.58}$$

By iterating the equation above forward and using the expressions for the net prices and the implementability constraint, we can obtain the transversality condition for assets (2.3). In order to prove that the allocation satisfies the government budget constraint, we subtract the individual budget constraint (2.2) from the aggregate resource constraint (2.13) and use the linear homogeneity of the production function (constant returns to scale):

$$a_{t+1} - k_{t+1} = (1 + r_t)(a_t - k_t) + g_t - \tau_t^K r_t a_t - \tau_t^L w_t l_t, \quad t > 0. \tag{2.59}$$

If we denote $d_t \equiv a_t - k_t$, we obtain both the condition for capital-market clearing (2.14) and the government budget constraint (2.8), thus proving that the implementation holds in both directions. □

2.B Proof of Lemma 2.2

Proof. When the agent's time horizon is finite and her preferences are not time-additive, the government's Lagrangian becomes:

$$\begin{aligned} \mathcal{L} = U(c_1, \ldots, c_T, l_1, \ldots, l_T) + \theta \left(\sum_{t=0}^{T} U_{c_t} c_t + U_{l_t} l_t \right) \\ + \sum_{t=0}^{T} \eta_t \left(f(k_t, l_t) - c_t - g_t - k_{t+1} + (1 - \delta) k_t \right), \end{aligned} \tag{2.60}$$

where θ is the multiplier on the implementability constraint and η_t is the multiplier on the period t aggregate resource constraint.

We can define the general equilibrium elasticities H_t^c and H_t^l, which are the equivalent of the general equilibrium elasticities ε_t^c and ε_t^l in a setting without time additivity:

$$H_t^c \equiv -\frac{\sum_{i=0}^{T} \left(U_{c_i c_t} c_i + U_{l_i c_t} l_i \right)}{U_{c_t}}, \quad H_t^l \equiv -\frac{\sum_{i=0}^{T} \left(U_{c_i l_t} c_i + U_{l_i l_t} l_i \right)}{U_{l_t}}, \quad t > 0. \tag{2.61}$$

This allows us to rewrite the government's first-order conditions with respect to consumption c_t, labour l_t and capital k_{t+1} as, respectively:

$$(1 + \theta) U_{c_t} - \theta U_{c_t} H_t^c = \eta_t, \quad t > 0, \tag{2.62}$$

$$(1 + \theta) U_{l_t} - \theta U_{l_t} H_t^l = -\eta_t f_{l_t}, \quad t > 0, \tag{2.63}$$

$$\frac{\eta_t}{\eta_{t+1}} = 1 + f_{k_{t+1}} - \delta = 1 + r_{t+1}, \quad t > 0. \tag{2.64}$$

Using the weak separability of U, we can rewrite the marginal rate of substitution between present and future consumption as:

$$\frac{U_{c_t}}{U_{c_{t+1}}} = \frac{h_{c_t}}{h_{c_{t+1}}}, \quad t > 0. \tag{2.65}$$

Since h is homothetic, the ratio $\frac{h_{c_t}}{h_{c_{t+1}}}$ is a function Ω of c_t and c_{t+1} only:

$$\frac{h_{c_t}}{h_{c_{t+1}}} = \Omega\left(\frac{c_t}{c_{t+1}}\right), \quad t > 0. \tag{2.66}$$

This allows us to rewrite the Euler equation (2.7) as:

$$\frac{c_t}{c_{t+1}} = \Omega^{-1}\left(R_{t+1}\right), \quad R_{t+1} \equiv 1 + (1-\tau^K_{t+1})r_{t+1}, \quad t > 0. \tag{2.67}$$

This expression shows that the consumption goods are linearly related in every period:

$$c_{t+1} = \gamma^C_{t+1} c_t, \quad t > 0, \tag{2.68}$$

where γ^C_{t+1} is a constant. This suggests that at the optimum, the entire vector of consumptions $(c_0, \ldots, c_t, \ldots, c_T)$ can be expressed as:

$$(c_0, \ldots, c_s, \ldots, c_T) = (\gamma^C_0, \ldots, \gamma^C_s, \ldots, \gamma^C_T)c_t, \quad t > 0, \tag{2.69}$$

where $\gamma^C_s = 1$ if $s = t$ for all t. Similarly, the entire vector of labour supplies can be expressed as:

$$(l_0, \ldots, l_s, \ldots, l_T) = (\gamma^L_0, \ldots, \gamma^L_s, \ldots, \gamma^L_T)l_t, \quad t > 0, \tag{2.70}$$

where $\gamma^L_s = 1$ if $s = t$. Using the expressions above to rewrite U, we can express the general equilibrium elasticities in a much simpler format, while taking into account that $\gamma^L_s = \gamma^C_s = 1$ if $t = s$:

$$H^c_t = -\frac{1}{U_{c_t}} \sum_{s=0}^{T} \left[\frac{1}{\gamma^C_s \gamma^C_t} c_s U_{c_t c_t} + \frac{1}{\gamma^L_s \gamma^C_t} l_s U_{c_t l_t}\right] = -\frac{1}{U_{c_t}} \sum_{s=0}^{T} \left[U_{c_t c_t} c_t + U_{c_t l_t} l_t\right], \quad t > 0, \tag{2.71}$$

$$H^c_{t+1} = -\frac{\gamma^C_{t+1}}{U_{c_t}} \sum_{s=0}^{T} \left[c_{t+1} \frac{1}{(\gamma^C_{t+1})^2} U_{c_t c_t} + \frac{1}{\gamma^C_{t+1}\gamma^L_{t+1}} U_{c_t l_t} l_{t+1}\right] = -\frac{1}{U_{c_t}} \sum_{s=0}^{T} \left[U_{c_t c_t} c_t + U_{c_t l_t} l_t\right]. \tag{2.72}$$

Equations (2.71) and (2.72) show that due to the properties of the utility function, the general equilibrium elasticity for consumption is constant, which suggests that $H_t^c = H_{t+1}^c$, for all t. Combining this result with the government's Euler equation (2.19) and the individual's Euler equation (2.7) gives the result $\tau_t^K = 0$ for all t. $\square$

2.C Proof Corrollary 2.3

Proof. We assume that utility is time separable, additively separable between consumption and leisure and homothetic in consumption:

$$U = \sum_{t=0}^{\infty} \beta^t \left(u(c_t) - v(l_t)\right), \tag{2.73}$$

where $u(c_t)$ is homothetic. By the separability of the utility function, we can rewrite the general-equilibrium elasticity ε_t^c as:

$$\varepsilon_t^c = \frac{u_{c_t c_t} c_t}{u_{c_t}}, \quad t > 0. \tag{2.74}$$

By homotheticity of u, we know that the following holds for any level of ϕ:

$$\frac{u_{c_t}(c_t)}{u_{c_{t+1}}(c_{t+1})} = \frac{u_{c_t}(\phi c_t)}{u_{c_{t+1}}(\phi c_{t+1})}, \quad t > 0. \tag{2.75}$$

Since the expression above can be treated as an identity, we can also differentiate it w.r.t ϕ and set ϕ to 1:

$$\frac{u_{c_t c_t} c_t}{u_{c_t}} = \frac{u_{c_{t+1} c_{t+1}} c_{t+1}}{u_{c_{t+1}}}, \quad t > 0. \tag{2.76}$$

Since the expression above holds for any t, we can conclude that ε_t^c is constant for all t. $\square$

2.D Proof of Lemma 2.3

Proof. We first prove that an implementable allocation satisfies the implementability constraint (2.15) and the aggregate resource constraint (2.50). Since an implementable allocation solves the individual's problem by definition, we can use the individual's optimality conditions, transversality condition for assets and the budget constraint to derive the implementability constraint. Furthermore, an implementable allocation satisfies the aggregate resource constraint by construction. Next, we prove that an allocation $\{c_t, l_t, g_t, x_{t+1}\}_{t=0}^{\infty}$ that satisfies (2.15) and (2.50) is implementable, given the factor prices r and w. The proof follows exactly the same steps as the one for the general equilibrium case in Appendix 2.A: we use the individual's first-order conditions to calculate the taxes $\{\tau_t^L, \tau_{t+1}^K\}_{t=0}^{\infty}$

that would implement the allocation. We then define the private asset path such that the individual budget constraint (2.2) holds and use the implementability constraint (2.15) and the private optimality conditions (2.6) and (2.7) to prove that the transversality condition for private assets (2.3) holds. The last step of the proof involves subtracting the aggregate resource constraint (2.50) from the individual's budget constraint (2.2) and defining the government debt $d_t = a_t - x_t$. This proves that both the condition for capital-market equilibrium (2.52) and the government budget constraint (2.8) hold. □

Chapter 3

Optimal Taxation of Capital Income with Heterogeneous Rates of Return*

*This Chapter is based on Gerritsen et al. (2019).

3.1 Introduction

Income inequality is rising in most parts of the world, in part due to rising inequality in capital income and wealth (Alvaredo et al., 2018). Piketty (2014)'s Capital in the Twenty-First Century brought the question of how governments should tax capital back to the center of the policy debate. Arguments against taxation of capital income date back to Mill (1848) and Pigou (1928), who argued that taxes on capital income amount to taxing labor income twice: first when it is earned, second when it is saved. If all inequality in capital income derives only from inequality in labor income, then it is of no independent concern for the optimal taxation of capital income. Taxes on capital income would redistribute income and distort labor supply in the same way as taxes on labor income, but they would also distort saving decisions. Hence, it would be better not to tax capital income at all (Atkinson and Stiglitz, 1976).[1]

The view that taxes on capital income are not helpful for income redistribution has been highly influential in academic and policy debates (Stiglitz, 2018). However, it critically hinges on the assumption that all individuals obtain the same rate of return on their savings, regardless of their earning abilities or wealth. This assumption has become untenable. A large and growing body of empirical evidence shows that people differ in their returns on savings, and that these returns are systematically

[1]The result that taxes on capital income are undesirable to redistribute income requires that individuals have identical and weakly separable preferences between consumption at different dates and leisure. Intuitively, conditional on labor income, all individuals then save the same amount (Atkinson and Stiglitz, 1976).

related to measures of ability and wealth. Importantly, these return differences are persistent and present even after controlling for risk-taking behavior. This evidence strongly suggests that inequality in capital income does not simply derive from inequality in labor income.

There are roughly three strands of empirical literature that speak to the importance of return heterogeneity. First, the most direct evidence simply documents differences in rates of return across the population. A seminal contribution is Yitzhaki (1987), who studies a subset of U.S. tax returns from 1973 and finds that rates of return increase with income. Piketty (2014) documents that universities with larger endowments are able to generate substantially larger returns on their investments than universities with smaller endowments. Saez and Zucman (2016) find the same for all registered US foundations. More recently, Fagereng et al. (2020) and Bach, Calvet, and Sodini (2018) find convincing evidence of significant return heterogeneity on the basis of administrative data on the populations of taxpayers in, respectively, Norway and Sweden over multiple years. Fagereng et al. (2020) find that differences in rates of return are important, persistent, and attributable to individual-specific factors that cannot be explained by observables, such as differences in the allocation of wealth between risky and safe assets. Moving from the 10th to the 90th percentile of the distribution of gross financial wealth (excluding non-financial wealth, such as businesses and housing), the average rate of return increases by 1.6 percentage points. This figure is only slightly lower if they restrict attention to safe assets or if they control for the volatility of the underlying portfolio.

Second, a large literature in finance documents that richer individuals tend to make fewer mistakes in their investments. An abundance of evidence shows that individuals do not optimally diversify their portfolios (e.g., Benartzi and Thaler, 2001; Choi, Laibson, and Madrian, 2005; Calvet, Campbell, and Sodini, 2007; Goetzmann and Kumar, 2008; Von Gaudecker, 2015). Furthermore, individuals consistently fail to optimize their financial portfolio even conditional on risk, for example by exposing themselves to excess interest and fee payments (Barber, Odean, and Zheng, 2005; Agarwal et al., 2009; Choi, Laibson, and Madrian, 2010, 2011). Investment mistakes may also be facilitated by fraudulent financial intermediaries that cater to financially unsophisticated clients (Egan, Matvos, and Seru, 2018). Unsurprisingly, investment mistakes are linked to individuals' financial literacy or sophistication, which itself is positively associated with education and wealth (e.g., Van Rooij, Lusardi, and Alessie, 2011; Lusardi and Mitchell, 2011; Lusardi, Michaud, and Mitchell, 2017). See also Campbell (2016) for a recent overview on mistakes in household finance. A natural implication of this evidence is that richer individuals obtain higher rates of return on their savings.

Third, recent research suggests that return heterogeneity is necessary to reconcile life-cycle models with observed patterns of wealth inequality. In particular, Benhabib, Bisin, and Luo (2019) and Gabaix et al. (2016) argue that return heterogeneity is needed to explain the dynamics of the fat, right tail of the U.S. wealth distribution. Importantly, Gabaix et al. (2016) emphasize both "type dependence" and "scale dependence" in return heterogeneity. That is, they argue that an individual's rate of return could depend on both his underlying type – e.g., his cognitive ability – and his level of wealth. Lusardi, Michaud, and Mitchell (2017) and Kacperczyk, Nosal, and Stevens (2018) emphasize the importance of heterogeneity in returns due to differences in financial sophistication to explain inequality in wealth and capital income. Indeed, Lusardi, Michaud, and Mitchell (2017) suggest that 30-40 percent of inequality in retirement wealth can be explained by return heterogeneity.

We derive the implications of return heterogeneity for optimal non-linear taxes on both labor and capital income. We study a two-period version of the Mirrlees (1971) model. Individuals differ in their ability and choose how much to work and how much to save in the first period of their life cycle. They consume all their savings and the returns on their savings in the second period of their life cycle.[2] Labor income is a function of labor supply and ability, whereas capital income is a reduced-form function of savings and ability, which is able to capture a number of microfoundations for return heterogeneity. The government can only observe labor income and capital income, and not ability. As a result, it must rely on distortionary taxes on labor and capital income to optimally redistribute income. We abstract from risk to focus our analysis solely on the implications of systematic heterogeneity in returns. We derive optimal taxes for two different microfoundations of return heterogeneity. In the first microfoundation, rates of return reflect type dependency as they are determined by ability. In the second microfoundation, rates of return reflect scale dependency and are determined by the amount of wealth. The two microfoundations generate two different reasons to tax capital income.

First, we consider type-dependent returns. Individuals have access to both a freely traded asset with a fixed rate of return, and a closely held asset with decreasing returns to capital and positive marginal returns to ability. Increasing returns to ability may reflect a positive association between earnings ability and entrepreneurial talent. In equilibrium, individuals equate the marginal rates of return of freely traded and closely held assets. While everyone therefore invests at the same marginal rates of return, individuals with higher ability obtain higher average rates of return. The resulting capital income increases with ability – for given savings. Capital income therefore reveals information about ability in addition to what is revealed by labor income. This implies that the government should tax both capital income and labor income at positive marginal rates to optimally redistribute income. The optimal tax on capital income trades off (additional) redistributional gains against distortions in saving.

[2] The life-cycle structure of our model is similar to that of Ordover and Phelps (1979), who also analyse optimal taxes on capital income in a two-period OLG version of the Mirrlees (1971) model. We assume that preferences over consumption and leisure are identical and separable, so that optimal taxes on capital income would be zero in the absence of return heterogeneity, see also Atkinson and Stiglitz (1976).

We derive an expression for the Pareto-efficient structure of taxes on capital income and labor income. The optimal dual tax structure equates the marginal distortions of both taxes to achieve the same amount of redistribution. The optimal dual tax structure does not depend on social welfare weights, but only on sufficient statistics with empirical counterparts: tax wedges and elasticities of labor income and capital income. The critical term in the optimal tax formula is the elasticity of capital income with respect to ability, conditional on labor income. The larger the degree of return heterogeneity, the larger this elasticity, and thus the larger the optimal tax on capital income. This elasticity, and therefore the optimal tax on capital income, is zero in the absence of return heterogeneity. Hence, our model nests Atkinson and Stiglitz (1976) as a special case. Furthermore, the optimal tax on capital income is decreasing in the elasticity of capital income with respect to the after-tax rate of return. The larger this elasticity, the more distortionary is the tax on capital income. Finally, the optimal tax on capital income is increasing in the tax wedge on labor income and the elasticity of labor income with respect to the net-of-tax rate on labor income. A large tax wedge on labor income indicates that taxes on labor income are more distortionary compared to taxes on capital income, for the same amount of income redistribution, hence capital income should be taxed relatively more.

Second, we consider scale-dependent returns. Individuals with more wealth obtain higher marginal rates of return due to positive scale effects. Implicitly, there is a failure of the capital market that prevents the poor from investing in assets with high returns. As a result, not all differences in marginal rates of return are arbitraged away. The failure of the capital market thus results in a misallocation of capital: everyone could be made better off if the rich would save on behalf of the poor. We show that a positive tax on capital income is optimal, because it alleviates the misallocation of capital. Intuitively, a positive tax on capital income redistributes from the rich to the poor in the second period. As individuals smooth consumption over time, the capital tax induces the rich to save relatively more and the poor to save relatively less. This increases the total amount of resources in the economy because the rich are able to obtain higher rates of return than the poor. We derive an intuitive ABC-style formula for Pareto efficient capital taxes in the spirit of Diamond (1998) and Saez (2001), where marginal returns to saving replace the standard welfare weights. The optimal marginal tax rate at a given level of capital income trades off the standard saving distortions against the benefits of an improved capital allocation. Optimal marginal tax rates are increasing in the degree of return heterogeneity. We also derive a simple formula for the optimal top tax rate on capital income.

Finally, we numerically simulate our model to obtain a quantitative sense of how important heterogeneity in capital returns is for optimal tax policy. We focus our numerical simulations on the first case, where heterogeneous returns originate from ability differences and not from differences in saving. We calibrate our model on US data on the distribution of income, but model return heterogeneity by using Norwegian estimates from Fagereng et al. (2020). We find that optimal taxes on capital income are substantial in all our simulations. In our baseline simulation, optimal taxes on capital income are on average around 12 percent, while optimal top rates on capital income are around 25 percent. However, we are not confident to make strong claims on the exact level of the optimal tax rate on capital income, except that it can be sizable. We demonstrate that optimal taxes on capital income critically depend on the elasticity of saving with respect to net returns and the extent of return heterogeneity. There is neither certainty nor consensus on these empirical objects. Moreover, the optimal non-linear marginal tax rates on capital income are generally increasing in capital income. We also provide an empirically reasonable calibration of the top tax rate in the case in which return heterogeneity stems from scale effects. We then find that Pareto optimal top rates on capital income can easily exceed 15 percent. Our calculations show that return heterogeneity may call for substantially positive marginal tax rates on capital income that are increasing in capital income.

The remainder of the paper is organized as follows. In Section 2, we discuss earlier results on optimal taxation of capital income and indicate how we contribute to this large literature. In Section 3, we introduce and discuss the theoretical setting of our paper. In Section 4, we explicitly show how our model is able to capture two different plausible microfoundations of return heterogeneity. In Section 5, we derive and discuss the optimal non-linear tax on capital income. Section 6 provides numerical simulations of optimal taxes on labor and capital income under realistic assumptions on return heterogeneity. A final section concludes. The Appendix contains some derivations and proofs of all propositions and lemmas.

3.2 Related literature

A number of key papers identify settings in which optimal taxes on capital income are zero (Atkinson and Stiglitz, 1976; Chamley, 1986; Judd, 1985). Much of the subsequent literature on capital taxation explores motives to tax capital income at positive rates. Taxes on capital income may be optimal to alleviate distortions of labor taxes on labor supply (Corlett and Hague, 1953; Atkinson and Stiglitz, 1976; Erosa and Gervais, 2002; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018), human capital (Jacobs and Bovenberg, 2010), or to contain tax shifting between labor and capital income (Christiansen and Tuomala, 2008; Reis, 2010). They may be an efficient instrument for redistribution if saving preferences are increasing with ability (Mirrlees, 1976; Saez, 2002; Diamond and Spinnewijn, 2011), if individuals have heterogeneous preferences for wealth itself (Saez and Stantcheva, 2018)[3]

[3] Saez and Stantcheva (2018) argue that taxing capital income could be desirable if individuals derive utility from wealth. They extend their results to the case where returns are heterogeneous. However, it not clear whether return heterogeneity increases or decreases the optimal tax on capital income, or even whether it affects the optimal tax at all.

or if endowments and inheritances positively correlate with labor income but cannot be directly taxed (Cremer, Pestieau, and Rochet, 2001). In models with overlapping generations, taxes (or subsidies) on capital income may be helpful to correct dynamic inefficiencies in capital accumulation (Ordover and Phelps, 1979; Atkinson and Sandmo, 1980; King, 1980). Finally, there are papers that derive optimally positive taxes on capital income in settings with idiosyncratic shocks to labor productivity and missing capital or insurance markets (Aiyagari, 1995; Hubbard and Judd, 1986; Diamond and Mirrlees, 1978; Golosov, Kocherlakota, and Tsyvinski, 2003; Conesa, Kitao, and Krueger, 2009; Jacobs and Schindler, 2012).

Our paper is most closely related to a small number of papers that also study optimal taxation with heterogeneous returns to capital. Stiglitz (1985, 2000, 2018) has conjectured but not formally shown that optimal taxes on capital income are positive if rates of return depend on ability. We confirm this conjecture. Gahvari and Micheletto (2016) and Kristjánsson (2016) study the two-type optimal tax framework of Stiglitz (1982), and show that optimal taxes on capital income are positive if rates of return are higher for the high-ability type. We contribute to these papers in a number of ways. First, we show that optimal taxes on capital income are positive if rates of return are an increasing function of wealth itself, i.e., if returns are scale dependent rather than type dependent. The government can then improve the allocation of capital by reducing taxes on labor income in the first period and raising taxes on capital income in the second period. In contrast, Gahvari and Micheletto (2016) and Kristjánsson (2016) mistakenly conclude that scale dependency of returns does not provide a reason to tax capital income, since they assume that all taxes are levied in the same period. Second, we study an economy with a continuum of types as in Mirrlees (1971). This allows us to derive meaningful optimal tax formulas in terms of sufficient statistics, as well as gain more insight into the shape of the optimal non-linear tax on capital income. We show that optimal taxes on capital income are increasing with capital incomes. Third, we derive conditions for the Pareto-efficient structure of taxes on capital income and labor income that does not depend on social welfare weights. Fourth, we provide numerical simulations of optimal non-linear taxes on capital income.

Guvenen et al. (2019) numerically study optimal linear tax systems in a macroeconomic model with overlapping generations that differ in their returns to capital. The way they model return heterogeneity is similar to our first microfoundation with common returns to a freely traded asset and type-dependent returns to a closely-held asset. Moreover, they allow for borrowing constraints. In the absence of borrowing constraints, they find that optimal taxes on capital income are positive. This confirms our findings in the first microfoundation: we both find an optimal dual tax structure that smooths out distortions over the labor and capital tax bases. However, if borrowing constraints are binding, Guvenen et al. (2019) find that individuals are not able to equate the marginal returns on the freely traded and closely-held assets. This leads to a misallocation of capital in which constrained individuals obtain a higher rate of return than unconstrained individuals. Comparable to our second microfoundation with scale-dependent returns, the government then wants to use the tax system to redistribute from low- to high-return individuals. In the case of Guvenen et al. (2019), this is done by setting lower and even negative taxes on capital income, and positive taxes on wealth. Intuitively, constrained individuals with high rates of return are more likely to have a high level of capital income. Thus, the government can redistribute towards high-return individuals by reducing taxes on capital income and increasing taxes on wealth. In contrast, we find that misallocation of capital provides a reason for higher taxes on capital income. Intuitively, the government wants to redistribute from poor low-return to rich high-return individuals before investments take place and revert this redistribution, once returns are realized. This is achieved by reducing marginal taxes on (early-life) labor income and increasing marginal taxes on (later-life) capital income.

Finally, there are papers that focus on optimal taxation if returns differ due to risk and portfolio choice. Varian (1980) shows that optimal taxes on capital income are positive if returns to savings feature idiosyncratic risk, and tax revenues can be returned in lump-sum fashion. Intuitively, taxes on capital income then provide social insurance by redistributing capital income from the lucky to the unlucky. Similarly, Gordon (1985) studies optimal taxation of capital income if individuals invest in both risk-free and risky assets. They find that taxes on capital income yield no insurance gains if there is only aggregate risk in capital returns and revenues are used to finance state-contingent lump-sum transfers. However, Christiansen (1993) and Schindler (2008) show that capital taxes are still optimal if the revenue is used to finance public goods. The optimal tax on capital income then balances the risk in private consumption against the risk in public good provision. Spiritus and Boadway (2017) study optimal taxes on both normal and above-normal returns to capital. Taxing above-normal returns insures idiosyncratic risk in capital income. A tax on the normal rate of return might be desirable if this leads to portfolio reallocation towards assets with idiosyncratic risk. All these studies suggest that the case for a positive optimal tax on capital income would be strengthened if we would allow for risky returns and portfolio choice.

3.3 Model

3.3.1 Individual behavior

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

$$c_1^n = z^n - T^n - a^n. \tag{3.1}$$

Savings yield capital income y^n, which is deterministic, and depends on the amount of savings and, potentially, on individual ability: $y^n = y(a^n, n)$. As we show later, this formulation allows us to capture plausible microfoundations of return heterogeneity related to closely-held businesses and scale economies in wealth investment. The case where returns on the assets from a closely-held business are increasing in the owner's ability could be captured by $y_n > 0$. The idea that returns are increasing in the total wealth of an individual could be captured by $y_{aa} > 0$. In the latter case, individuals generally differ in their marginal rate of return y_a. Thus, we implicitly allow for capital-market failures, such that differences in marginal rates of return are not necessarily arbitraged away. Taxes on capital income are denoted by τ^n, and second-period consumption equals the sum of savings and after-tax capital income:

$$c_2^n = a^n + y(a^n, n) - \tau^n. \tag{3.2}$$

T^n is a non-linear tax function of labor income z^n, and τ^n is a non-linear tax function of capital income y^n. We parameterize the tax schedules in a way that allows us to study the effects of exogenous shifts in their slopes and intercepts. This later helps us define behavioral elasticities and social welfare weights.[4] We write the tax schedules as the following functions:

$$T^n = T(z^n, \rho^T, \sigma^T) \equiv \tilde{T}(z^n) + \rho^T + \sigma^T z^n, \tag{3.3}$$

$$\tau^n = \tau(y^n, \rho^\tau, \sigma^\tau) \equiv \tilde{\tau}(y^n) + \rho^\tau + \sigma^\tau y^n, \tag{3.4}$$

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

[4] It is not uncommon to parameterize non-linear tax schedules to derive the comparative statics, see, e.g., Christiansen (1981); Immervoll et al. (2007); Jacquet, Lehmann, and Van der Linden (2013); Gerritsen (2016).

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

$$U^n = u(c_1^n, c_2^n) - v(z^n/n). \tag{3.5}$$

Utility of consumption $u(\cdot)$ is increasing, concave, and twice continuously differentiable. Disutility of work $v(\cdot)$ is increasing, strictly convex and twice continuously differentiable. Utility is separable between consumption and labor supply, so there is no reason to tax capital income in the absence of return heterogeneity (Atkinson and Stiglitz, 1976). Substituting first- and second-period consumption and the parameterized tax schedules into the utility function and optimizing over savings and labor income yields the following first-order conditions:

$$\frac{v'(z^n/n)}{u_1(c_1^n, c_2^n)} = (1 - T'(z^n, \rho^T, \sigma^T))n, \tag{3.6}$$

$$\frac{u_2(c_1^n, c_2^n)}{u_1(c_1^n, c_2^n)} = \frac{1}{1 + (1 - \tau'(y(a^n, n), \rho^\tau, \sigma^\tau))y_a(a^n, n)} \equiv \frac{1}{R^n}. \tag{3.7}$$

We denote marginal tax rates by a prime: $T'(z^n, \rho^T, \sigma^T) \equiv \partial T(z^n, \rho^T, \sigma^T)/\partial z^n$ and $\tau'(y^n, \rho^\tau, \sigma^\tau) \equiv \partial \tau(y^n, \rho^\tau, \sigma^\tau)/\partial y^n$. Other partial derivatives are denoted by a subscript. Thus, $u_1(\cdot)$ and $u_2(\cdot)$ are the marginal utility of first- and second-period consumption, and $y_a(\cdot)$ denotes the marginal rate of return. Eq. (3.6) shows that the marginal rate of substitution between (first-period) consumption and leisure must equal the marginal after-tax wage rate. Eq. (3.7) shows that the marginal rate of substitution between first- and second-period consumption must equal the individual's discount factor. We define the inverse of the discount factor – or one plus the after-tax rate of return – as $R^n \equiv 1 + (1 - \tau')y_a$.[5]

We impose a number of assumptions that help us derive the optimal non-linear tax schedules. First, we require that both tax schedules are twice continuously differentiable. This ensures that the individual first-order conditions are differentiable. Second, we assume that second-order conditions are satisfied, and that eqs. (3.6) and (3.7) describe a unique and global maximum for utility. This guarantees that individual behavior is differentiable and thus that marginal changes in taxes lead to marginal responses in earnings. It also implies that the equilibrium values of both tax bases y^n and z^n are monotonically increasing in ability n. These assumptions correspond to Assumption 2 in Jacquet and Lehmann (2017).

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

3.3.2 Behavioral elasticities

Behavioral elasticities of the tax bases play an important role in the optimal tax expressions that we derive below. To define these elasticities, we first write the tax bases as functions of the tax parameters. The first-order condition for savings in eq. (3.7), together with the definitions of first- and second-period consumption (3.1) and (3.2), implicitly determines equilibrium savings as a function of labor income, tax parameters, and ability. This allows us to write equilibrium savings as $a^n = \tilde{a}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$, where the superscript c indicates conditionality on labor income z^n. Since capital income is a function of savings and ability, we can write equilibrium capital income as a function of the same arguments $y^n = \tilde{y}^c(z^n, \rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$. Both first-order conditions in eqs. (3.6) and (3.7), together with the definitions of first- and second-period consumption, determine labor income as a function of tax parameters and ability. This allows us to write equilibrium labor income as $z^n = \tilde{z}(\rho^T, \rho^\tau, \sigma^T, \sigma^\tau, n)$.[6]

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

$$e_z^n \equiv -\left(\frac{\partial \tilde{z}}{\partial \sigma^T} - z^n \frac{\partial \tilde{z}}{\partial \rho^T}\right) \frac{1-T'}{z^n}. \tag{3.8}$$

The elasticity in eq. (3.8) measures the percentage change in labor income if the net-of-tax rate $1-T'$ is exogenously raised by one percent, while utility is kept constant. It captures the total impact on labor income, taking into account the effect a change in earnings can have on the marginal tax rate if the tax function is non-linear.[7] The term within brackets gives the Slutsky decomposition of the compensated response in labor income to an increase in marginal taxes.[8]

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

$$e_{y|z}^n \equiv -\frac{1}{y_a}\left(\frac{\partial \tilde{y}^c}{\partial \sigma^\tau} - y^n \frac{\partial \tilde{y}^c}{\partial \rho^\tau}\right) \frac{R^n}{y^n}. \tag{3.9}$$

The elasticity in eq. (3.9) measures the percentage change in capital income if the after-tax rate of return $R^n = 1 + (1-\tau')y_a$ is exogenously raised by one percent, while utility is kept constant. It captures the total impact on capital income, taking into account the effect a change in capital income can have on the marginal tax rate if the tax function is non-linear. Again, the term within brackets is the Slutsky decomposition of the compensated response of capital income to an increase in the

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

[7] In terms of Jacquet and Lehmann (2017), e_z^n is a 'total elasticity' rather than a 'direct elasticity.' The elasticity measures the effect on labor income of a given change in the tax parameters σ^T and ρ^T rather than a given change in the marginal tax rate $T'(z, \rho^T, \sigma^T)$. Total elasticities are also used by, e.g., Jacquet, Lehmann, and Van der Linden (2013), Jacobs and Boadway (2014), Gerritsen (2016), and Scheuer and Werning (2017).

[8] We formally derive the Slutsky decompositions in Appendix 3.B.1.

marginal tax rate. Furthermore, $e^n_{y|z}$ is a conditional elasticity, in that it measures the behavioral change in capital income while holding labor income constant.

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

$$\xi^n_z \equiv \frac{\partial \tilde{z}}{\partial n}\frac{n}{z^n}, \qquad \xi^n_{y|z} \equiv \left.\frac{\partial \tilde{y}}{\partial n}\right|_z \frac{n}{y^n}. \tag{3.10}$$

The first elasticity ξ^n_z measures the percentage change in labor income due to a one percent increase in ability. The second elasticity $\xi^n_{y|z}$ measures the percentage change in capital income due to a one percent increase in ability, while holding labor income constant.

3.4 Two microfoundations of return heterogeneity

3.4.1 Type-dependent returns: entrepreneurial investments

It is instructive to consider two plausible microfoundations for capital income $y(a^n, n)$ that could generate heterogeneity in rates of return. These two different microfoundations loosely correspond to what Gabaix et al. (2016) call type-dependent and scale-dependent returns. We first consider type-dependent returns. In particular, we consider an economy in which individuals can invest in two different types of assets. They can invest in a closely held asset that is specific to their type and could be interpreted as entrepreneurial investment. And they can invest in an asset that is freely traded in capital markets.

Individual n invests b^n in the closely held asset. This yields a total return that is a function of invested capital and ability: $\pi^n = \pi(b^n, n)$. The closely held asset exhibits decreasing returns to capital ($\pi_b > 0$ and $\pi_{bb} < 0$) and increasing returns to ability ($\pi_n > 0$). The latter assumption reflects the idea that high ability helps to find and select successful business ventures. Individual n invests the remainder of his savings, $a^n - b^n$, in the freely traded asset, which yield a constant rate of return r that is common for all individuals.

Capital income is now given by:

$$y^n = r(a^n - b^n) + \pi(b^n, n). \tag{3.11}$$

Individuals allocate their savings over the two assets in a way that maximizes their capital income (provided that the marginal tax on capital income is below 100 percent, $\tau' < 1$). Maximizing y^n in eq. (3.11) with respect to b^n yields $\pi_b(b^n, n) = r$. Thus, individuals invest in the closely-held asset up to the point at which its marginal return equals that on the commonly traded asset. This implicitly determines entrepreneurial investment as a function of ability alone: $b^n = b(n)$. Substituting this into the equation for capital income (3.11) yields:

$$y^n = y(a^n, n) = ra^n + \pi(b(n), n) - rb(n). \tag{3.12}$$

Hence, the general formulation $y^n = y(a^n, n)$ can capture the special case of entrepreneurial investments. Under this microfoundation, capital income is linear in savings and increasing in ability: $y_a = r$ and $y_n > 0$.[9]

3.4.2 Scale-dependent returns: scale economies in wealth investment

The second microfoundation of capital income $y(a^n, n)$ relies on scale economies in wealth investment. Scale economies may originate from the fixed costs associated with raising rates of return. For example, an individual needs a savings account with a bank to earn any interest on savings at all. Because banks typically charge their account holders fixed periodic fees, it only makes sense to open an account and obtain a positive rate of return if savings are large enough to cover these fixed fees. Moreover, to participate in higher-yielding assets such as equity, one needs to invest in at least some basic financial knowledge or acquire the costly services of a wealth manager. Again, it only makes sense to pay for these higher yields if the invested wealth is sufficiently large. As a consequence, individuals with more wealth are likely to obtain higher rates of return.

We can capture scale effects in our model by assuming that individuals invest x^n of their savings to raise the returns on the remainder of their savings. These investments consist of search costs, fixed fees, and the costs of obtaining financial know-how. This leaves an amount $a^n - x^n$ to be invested at a rate of return $r(x^n) \geq 0$ with $r'(x^n) \geq 0$. We assume that all investment costs are deductible from the capital tax base but not separately observed by the government and thus not separately taxed.[10] Taxable capital income is then given by:

$$y^n = r(x^n)(a^n - x^n) - x^n. \tag{3.13}$$

Individuals invest in financial services to maximize their capital income. Maximizing y^n in eq. (3.13) with respect to x^n yields $r'(x^n)(a^n - x^n) = 1 + r(x^n)$.[11] The left-hand side gives the gains from

[9] Both follow from the partial derivatives of $y(a, n)$ in eq. (3.12). $y_a = r$ follows trivially. Application of the envelope theorem yields $y_n = \pi_n \geq 0$.

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

[11] The second-order condition that ensures an interior solution is given by $r''(x^n)(a^n - x^n) < 2r'(x^n)$. It is intuitively plausible that there is an upper limit to which the rate of return can rise by investing more and more in search costs, wealth management fees, and financial know-how. If the rate of return $r(x)$ is indeed increasing in x at a decreasing rate, then the second-order condition is always satisfied.

investing one more unit of resources in obtaining a higher rate of return. The right-hand side denotes the opportunity costs of doing so. The equilibrium condition implicitly determines investment costs as a function of savings $x^n = x(a^n)$, with $x'(a^n) \geq 0$. Intuitively, the larger one's wealth, the stronger are the incentives to increase the rate of return. Substituting this into the expression for capital income yields:

$$y^n = y(a^n, n) = r(x(a^n))(a^n - x(a^n)) - x(a^n). \tag{3.14}$$

Hence, the general formulation $y^n = y(a^n, n)$ also captures scale economies in wealth investment. In that case, capital income is convex in savings and does not (directly) depend on ability: $y_a \geq 0$, $y_{aa} \geq 0$, and $y_n = 0$.[12]

Individuals with different levels of wealth face different marginal rates of return and therefore different marginal rates of transformation between first- and second-period consumption. The costs x can be interpreted as the costs of entering a specific financial market in which assets yield a rate of return $r(x)$. Thus, individuals with different levels of wealth effectively invest in segmented financial markets. As a result, there is not one single financial market that equates marginal rates of transformation. This means that there are potential Pareto-improving trades in the capital market that do not materialize. To see this, imagine that a high-wealth, high-return individual invests funds on behalf of a low-wealth, low-return individual at some intermediate interest rate. Such a transaction would be mutually beneficial because the low-wealth individual could gain access to the returns of higher-yielding assets, while the high-wealth individual would obtain even higher returns due to scale effects. Thus, implicit in the microfoundation is a market failure that keeps low-wealth individuals from accessing the higher-yielding investment opportunities of the rich.

3.5 Optimal taxation

3.5.1 Instrument set

We assume that the government can observe labor income and capital income at the individual level. Hence, the government can implement non-linear taxes on labor income and capital income. We rule out taxes on consumption and wealth for both theoretical and practical reasons. Theoretically, allowing for either consumption or wealth taxes would enable the government to tax away all heterogeneous returns with zero distortions, which we like to avoid.[13] In practice, 100 percent taxation of excess returns via wealth or consumption taxes will surely result in tax avoidance and evasion

[12]This follows from the partial derivatives of eq. (3.14). $y_n = 0$ follows trivially. Application of the envelope theorem yields $y_a = r(x(a^n)) \geq 0$ and, hence, $y_{aa} = r'x' \geq 0$.

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

due to cross-border shopping, international mobility of capital, and reduced entrepreneurial efforts, which are behavioral margins from which we like to abstract. Hence, our model captures the main policy trade-off for the optimal taxation of capital income with heterogeneous returns, while avoiding complexities with modeling cross-border shopping, capital mobility or entrepreneurial effort.

We also assume that capital income and labor income are taxed separately. In doing so we follow most of the literature, see e.g., see Saez and Stantcheva (2018). As a result, the marginal tax rate on one tax base does not depend on the size of the other tax base.[14]

3.5.2 Social welfare and government budget constraints

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

$$\mathcal{W} = \int_0^\infty W(U^n) f(n) \mathrm{d}n, \quad W'(U^n) > 0 \quad W''(U^n) \leq 0. \tag{3.15}$$

Social preferences for income redistribution are captured by concavity of either the welfare function W or the utility function U^n.

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

$$\mathcal{B}_1 = \int_0^\infty T(z^n, \rho^T, \sigma^T) f(n) \mathrm{d}n - g_1 = 0, \tag{3.16}$$

$$\mathcal{B}_2 = \int_0^\infty \tau(y^n, \rho^\tau, \sigma^\tau) f(n) \mathrm{d}n - g_2 = 0, \tag{3.17}$$

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

Instead of assuming exogenously fixed government assets, we could alternatively assume that the government has access to the same investment technology as individuals. In the first microfoundation with commonly traded assets, this would imply that the government could borrow and lend at the same, constant marginal rate of return as every individual. As a result, government debt would be completely neutral. The optimal net asset position of the government then becomes indeterminate due

[14]Renes and Zoutman (2014) show that separable tax schedules are sufficient to implement the full second-best optimum if there are no market failures, the government has a welfarist objective, and the population is characterized by one-dimensional heterogeneity. These assumptions are fulfilled in our first microfoundation with type-dependent returns. However, these assumptions are no longer fulfilled in our second microfoundation with scale-dependent returns due to the implicit failure of the capital market. Hence, separate tax schedules on labor and capital income may not generally implement the full second-best optimum.

to Ricardian equivalence.[15] Thus, assuming that the government's net asset position is exogenously fixed is innocuous in case of the first microfoundation.

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

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

$$\mathcal{L} = \frac{1}{\lambda_1}\mathcal{W} + \mathcal{B}_1 + \frac{1}{\lambda_1/\lambda_2}\mathcal{B}_2. \tag{3.18}$$

The government discounts future tax revenue at a rate λ_1/λ_2.

3.5.3 Excess burdens and social welfare weights

The optimal tax structure depends on the excess burdens and distributional benefits of taxation. We define the marginal excess burden as the revenue loss caused by a compensated increase in a marginal tax rate. The marginal excess burdens of taxes on labor and capital income for individual n are given by:

$$E_T^n \equiv -T'\left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}\right) - \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}\right), \tag{3.19}$$

$$E_\tau^n \equiv -T'\left(\frac{\mathrm{d}z^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau}\right) - \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}y^n}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau}\right). \tag{3.20}$$

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

An increase in marginal taxes potentially affects both tax bases, thereby affecting both first- and second-period revenue. Eq. (3.19) gives the marginal excess burden of the tax on labor income. The first term equals the revenue loss from a compensated response in labor income, and the second term equals the revenue loss from a compensated response in capital income. Eq. (3.20) gives the marginal excess burden of the tax on capital income. Again, the equation gives the revenue losses from compensated responses in both labor and capital income.

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

$$\alpha_1^n \equiv \frac{W'(U^n)u_1}{\lambda_1} - T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} - \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}, \tag{3.21}$$

$$\alpha_2^n \equiv \frac{W'(U^n)u_2}{\lambda_1} - T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau} - \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau}. \tag{3.22}$$

The social welfare weights consist of the (monetized) utility gains of providing individual n with an additional unit of income in period 1 or 2, and the change in revenue due to the income effects on both tax bases.

3.5.4 Optimal tax schedules

We solve for the optimal non-linear taxes on labor and capital income by using the tax-perturbation approach, which was pioneered by Saez (2001), and has more recently been extended and amended by Golosov, Tsyvinski, and Werquin (2014), Gerritsen (2016), and Lehmann et al. (2018). In particular, we consider small perturbations in the tax schedules on labor and capital income. In the optimum, such tax perturbations should have no effect on social welfare. Denote the density of labor income by $h(z)$ and the density of capital income by $g(y)$. The following Lemma presents optimality conditions for marginal taxes on labor and capital income.

Lemma 3.1. *In the tax optimum, the following two conditions characterize the optimal marginal tax rates on labor and capital income for all income levels z^n and y^n:*

$$E_T^n h(z^n) = \int_{z^n}^{\infty} (1-\alpha_1^m) h(z^m)\mathrm{d}z^m, \tag{3.23}$$

$$E_\tau^n g(y^n) = \int_{y^n}^{\infty} \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^m\right) g(y^m)\mathrm{d}y^m. \tag{3.24}$$

Proof. See Appendix 3.A. □

The conditions in Lemma 3.1 are intuitively straightforward. Consider a small change to the marginal tax rate over a small interval of either tax base. The marginal deadweight cost of raising a marginal tax rate should equal its marginal redistributional gains. A small increase of the marginal tax rate on labor income around z^n distorts labor supply for all individuals with income around z^n. The excess burden associated with this distortion is given by the left-hand side of eq. (3.23). The perturbation also raises tax revenue from individuals who earn more than z^n. The redistributional gains of this are given by the right-hand side of eq. (3.23). In the same spirit, the left-hand side of eq. (3.24) gives the marginal excess burden of raising the marginal tax rate on capital income around y^n, and the right-hand side gives its redistributional gains.

3.5.5 Optimal taxation of labor income

Lemma 3.1 implicitly expresses optimal tax schedules in terms of marginal excess burdens and redistributional gains of taxation. To gain more insight into the shape of the optimal tax schedules, we explicitly write them in terms of wedges, elasticities, the income distribution, and social welfare weights. The following Proposition establishes the optimal tax wedge on labor income. For notational convenience, we suppress the tax parameters from the function arguments of the tax schedules, so that marginal tax rates at income levels z and y are written as $T'(z)$ and $\tau'(y)$. Moreover, we suppress the superscripts n in view of the perfect mapping between ability and labor and capital income.

Proposition 3.1. *The optimal tax wedge on labor income for all levels of labor income z is given by:*

$$\frac{T'(z)}{1-T'(z)}+\frac{sy_a\tau'(y)}{\lambda_1/\lambda_2}=\frac{1}{e_z}\frac{1-H(z)}{zh(z)}\Big(1-\bar{\alpha}_1^+(z)\Big), \tag{3.25}$$

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

Proof. Recall that $z^n = \tilde{z}(\rho^T,\rho^\tau,\sigma^T,\sigma^\tau,n)$. Thus, we can write $\mathrm{d}z^n/\mathrm{d}\sigma^T = \partial\tilde{z}/\partial\sigma^T$ and $\mathrm{d}z^n/\mathrm{d}\rho^T = \partial\tilde{z}/\partial\rho^T$. Second, note that $y^n = y(a^n,n)$ and $a^n = \tilde{a}^c(z^n,\rho^T,\rho^\tau,\sigma^T,\sigma^\tau,n)$. Moreover, a compensated change in the marginal tax rate on labor income only affects savings through labor income, i.e.: $\partial\tilde{a}^c/\partial\sigma^T - z^n\partial\tilde{a}^c/\partial\rho^T = 0$. We can therefore write $\mathrm{d}y^n/\mathrm{d}\sigma^T - z^n\mathrm{d}y^n/\mathrm{d}\rho^T = y_a\partial\tilde{a}^c/\partial z^n(\partial\tilde{z}/\partial\sigma^T - z^n\partial\tilde{z}/\partial\rho^T) = (1-T')y_a s^n(\partial\tilde{z}/\partial\sigma^T - z^n\partial\tilde{z}/\partial\rho^T)$. Substitute these expressions and the elasticity of eq. (3.8) into the definition of the excess burden in eq. (3.19), and substitute this into the optimality condition in eq. (3.23). Rearranging yields eq. (3.25). □

The left-hand side of eq. (3.25) gives the tax wedge on labor income for an individual with income z. To see this, consider a unit increase in after-tax labor income. This implies a $1/(1-T')$ increase in gross labor income, which leads to a revenue gain of $T'/(1-T')$. Moreover, it raises savings by s and capital income by $y_a s$, yielding a second-period revenue gain of $y_a s\tau'$, which the government discounts at a rate λ_1/λ_2. The right-hand side of eq. (3.25) is the standard expression for the optimal tax wedge on labor, see also Mirrlees (1971), Diamond (1998), and Saez (2001). The optimal tax wedge on labor income is decreasing in the elasticity of labor income at z, e_z, the relative hazard rate of the income distribution at z, $zh(z)/(1-H(z))$, and the average of the social-welfare weights of individuals who earn more than z, $\bar{\alpha}_1^+(z)$. The only material difference with, e.g., Saez (2001), is that the tax wedge on labor income contains not only the tax on labor income, but also the tax on capital income. This is because a reduction in labor income causes individuals to save less, thereby lowering revenue from taxes on both labor and capital income. If the marginal propensity to save is zero ($s = 0$), reductions in labor income do not reduce future consumption, so that the standard Saez-formula results.

3.5.6 Optimal taxation of capital income

In this subsection, we present and discuss our main theoretical results: the expressions for the optimal tax on capital income in the presence of return heterogeneity. We first discuss the case of type-dependent returns, in which return heterogeneity originates from closely held assets. We then discuss the case of scale-dependent returns, in which return heterogeneity originates from scale economies in wealth investment. We end with a more general formulation of the optimal tax on capital income that captures both microfoundations as special cases.

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

Recall from Section 3.4 that one possible microfoundation for capital income $y(a,n)$ reflects the existence of closely held assets with decreasing returns to investments and increasing returns to ability. Together with a commonly traded asset, this microfoundation ensures that capital income is linear in savings and increasing in ability: $y_a = r$ and $y_n \geq 0$. The following Proposition then establishes the optimal tax on capital income as a function of the tax on labor income.

Proposition 3.2. *If capital income is linear in savings but increasing in ability ($y_a = r$ and $y_n \geq 0$ for all individuals), the optimal marginal tax rate on capital income for every level of capital income*

y is implied by:

$$\left(\frac{y_a \tau'(y)}{1+y_a}\right) e_{y|z} = \left(\frac{T'(z)}{1-T'(z)} + \frac{s y_a \tau'(y)}{1+y_a}\right) e_z \left(\frac{\xi_{y|z}}{\xi_z}\right) \geq 0. \tag{3.26}$$

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

Proof. See Appendix 3.B. □

An increase in the marginal tax rate on capital income has three welfare-relevant effects: 1) It redistributes income from rich to poor, 2) It distorts labor supply, and 3) It distorts savings behavior. In the tax optimum, the marginal benefits of income redistribution should be equal to the marginal costs of distorting savings and labor supply. A tax on labor income also redistributes from rich to poor and distorts labor supply, but it leaves savings undistorted. Thus, a positive tax on capital income is only desirable if, for the same redistributional benefits, it distorts labor supply less than a tax on labor income. The optimal tax on capital income then equates the marginal costs of distorting savings to the marginal benefits of distorting labor supply less than a tax on labor income. This is shown in eq. (3.26).

The left-hand side of eq. (3.26) gives the deadweight loss of the savings distortion of a tax on capital income. It equals the tax wedge on savings, $y_a \tau'(y)/(1+y_a)$, multiplied by the conditional elasticity of capital income with respect to the after-tax interest rate $e_{y|z}$. The right-hand side gives the deadweight loss of the labor supply distortions of a tax on labor income relative to a distributionally equivalent tax on capital income. The two terms in brackets represent the tax wedge on labor income. It is multiplied by e_z to form the marginal deadweight loss of a tax on labor income. This is multiplied by the ratio of ability elasticities $\xi_{y|z}/\xi_z$, which captures the degree to which the tax on capital income distorts labor supply less than the tax on labor income for the same amount of income redistribution. Eq. (3.26) shows that the optimal tax on capital income is positive only if the ratio $\xi_{y|z}/\xi_z$ is positive. This is the case if and only if capital income is increasing in ability for given savings, such that $y_n(a,n) > 0$. Thus, capital income should be taxed in the presence of type-dependent returns.

Intuitively, the ratio $\xi_{y|z}/\xi_z$ captures the extent to which ability correlates more strongly with capital income than with labor income. Taxes on labor and capital income are second-best, because the government cannot observe and therefore cannot tax ability. Instead, the government wants to tax those tax bases that provide information on ability. If $\xi_{y|z}/\xi_z = 0$, then capital income provides the same information about ability as labor income. In that case, taxes on capital income generate the same distortions on labor supply for the same redistribution in income, and, in addition, distort savings decisions. A tax on capital income is then less desirable than a tax on labor income to redistribute income and should optimally be set to zero. Hence, our model nests Atkinson and Stiglitz (1976) as the special case where returns are equal for all individuals.

If $\xi_{y|z}/\xi_z > 0$, capital income reveals more information about ability than labor income. Starting from a situation without taxes on capital income, but with positive taxes on labor income, the capital tax generates fewer distortions than the tax on labor income for the same redistribution of income. The capital tax generates only a second-order welfare loss in saving, but allows for a first-order welfare gain in labor supply. As a result, in the full optimum, it is desirable to set a positive tax on capital income even if this distorts savings decisions.

The optimality condition in eq. (3.26) shows that the optimal marginal tax rate on capital income depends on a limited number of key statistics. First of all, it is increasing in the ratio of ability elasticities $\xi_{y|z}/\xi_z$. Second, the optimal tax rate on capital income is decreasing in the conditional elasticity of capital income with respect to the after-tax interest rate $e_{y|z}$. The larger this elasticity, the larger the savings distortions associated with marginal tax rates on capital income. Third, the optimal tax on capital income is increasing in the compensated elasticity of labor income with respect to the net-of-tax rate e_z. The larger this elasticity, the larger are labor supply distortions relative to savings distortions, and thus the more desirable is the tax on capital income relative to a tax on labor income. Fourth, and for the same reason, the optimal tax on capital income is increasing in the tax wedge on labor income. Thus, provided that elasticities are relatively constant over the income distribution, the shape of the optimal tax schedule on capital income tracks the shape of the optimal tax schedule on labor income.

Proposition 3.2 is closely related to a number of earlier contributions to the literature on the optimal taxation of commodities and of capital income. While Atkinson and Stiglitz (1976) have shown that the government should not use taxes on saving if preferences are homogeneous and separable between consumption and leisure, many subsequent studies have focused on the implications of non-separability and heterogeneity of preferences. Mirrlees (1976) notes that "commodity taxes should bear more heavily on the commodities high-n individuals have relatively strongest tastes for" – meaning that we should focus on "the way in which demands change for given income and labor supply when n changes." This finding is echoed in subsequent studies by Christiansen (1984), Saez (2002), Diamond and Spinnewijn (2011), and Jacobs and Boadway (2014).[16] Similar to these studies, we also

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

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

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

find that capital income should be taxed if it is increasing in ability for given labor income. Contrary to these earlier studies, we show that this argument does not rely on taste heterogeneity or non-separability in preferences. Instead, savings may increase in ability because of empirically plausible heterogeneity in rates of return. This implies that budget constraints rather than saving preferences depend on n for given labor income. In that respect, our findings are also closely related to Cremer, Pestieau, and Rochet (2001), who find that taxes on capital income are desirable if endowments – which are part of the budget constraint – are increasing with ability.

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

The second microfoundation for capital income $y(a, n)$ reflects the existence of economies of scale in wealth management: $y_{aa} \geq 0$ and $y_n = 0$. The following Proposition establishes the optimality of positive taxes on capital income if rates of return are increasing in savings.

Proposition 3.3. *If capital income is convex in savings, but not directly affected by ability ($y_{aa} \geq 0$ and $y_n = 0$ for all individuals), the optimal marginal tax rate on capital income for every level of capital income y is implied by:*

$$\frac{\tau'(y)y_a}{1+\bar{y}_a} = \frac{1}{e_{y|z}} \frac{1-G(y)}{yg(y)} \frac{\bar{y}_a^+(y) - \bar{y}_a}{1+\bar{y}_a} \geq 0, \tag{3.27}$$

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

Proof. See Appendix 3.B. □

The capital tax now has four welfare-relevant effects: 1) It redistributes income from rich to poor, 2) It distorts labor supply, 3) It distorts savings behavior, and 4) It reallocates capital from the poor to the rich. In the tax optimum, the marginal benefits from income redistribution and reallocation of capital should be equal to the marginal costs of distorting savings and labor supply.

In contrast to the first microfoundation, capital income does not provide any more information about ability than labor income. This is because capital income is a function of savings only and does not directly depend on ability ($y_n(a^n, n) = 0$). As a result, for the same amount of income redistribution, a tax on capital income generates the same labor supply distortion as a tax on labor income. This implies that the first two welfare-relevant effects of the capital tax cancel out if the tax on labor income is set optimally. This explains why the social welfare weights and the tax wedge on labor income are both absent from the optimal tax expression in eq. (3.27). Instead, the optimal tax on capital income equates the marginal costs of distorting savings with the marginal benefits of reallocating capital.

The left-hand side of eq. (3.27) is the tax wedge on savings for individuals with capital income y. In the optimum, this wedge equals the social marginal gain of capital reallocation on the right-hand side. To get a better understanding of the capital reallocation effect, consider an increase in the marginal tax on capital income at y. This raises the tax burden of individuals with capital income above y. Suppose that the government engineers a distributionally equivalent change in the labor tax that keeps all individuals on the same level of utility. If the labor tax is optimized, this tax change has no implications for social welfare. The combined tax reform thus shifts the tax liability of rich individuals from the first to the second period of their life cycle. To counteract this – and thus to maintain their preferred intertemporal consumption profile – these individuals will save relatively more.[17] At the same time, the government can close its budget by reducing a lump-sum tax in the second period and raising a lump-sum tax in the first period. Shifting the tax liability for all individuals from the second to the first period induces an offsetting reduction in savings. The positive savings response of the rich generates an average rate of return of $\bar{y}_a^+(y)$, whereas the negative savings response of the entire population foregoes an average rate of return of $\bar{y}_a < \bar{y}_a^+(y)$. Hence, a positive tax on capital income generates additional resources to society. In present value terms, this resource gain equals the discounted difference between the two average marginal rates of return: $(\bar{y}_a^+(y) - \bar{y}_a)/(1 + \bar{y}_a)$.

Heterogeneity in marginal rates of return implies an absence of a financial market that equates marginal rates of intertemporal transformation. Poor low-return individuals are unable to invest in the relatively high-yielding assets of wealthy high-return individuals. The government partially replaces the missing financial transactions between the wealthy and the poor by taxing capital income rather than labor income. By shifting the tax burden of the wealthy from the first period (i.e., by taxing labor income less) to the second period (by taxing capital income more), the wealthy will raise their savings relatively more. And by shifting the tax burden of the poor from the second period to the first period, the poor will reduce their savings relatively more. Thus, by setting a positive marginal tax on capital income, the government effectively forces the wealthy to save on behalf of the poor.

The optimal tax formula resembles the standard ABC-formula of Diamond (1998) and Saez (2001), where the difference between social welfare weights of the average individual (i.e., 1) and the welfare social weight of the average individual above y is replaced by the discounted difference between the discounted value of marginal returns, i.e., $\bar{y}_a^+(y)/(1 + \bar{y}_a)$, and the discounted value of the average returns, i.e., $(\bar{y}_a/(1 + \bar{y}_a)$ for all individuals above y. The larger the conditional elasticity of capital income, $e_{y|z}$, the larger the savings distortions, and thus the lower the optimal tax on capital

[17] This is formally shown in eq. (3.72) of Appendix 3.B.2.

income. The larger the inequality in rates of return, $(\bar{y}_a^+(y) - \bar{y}_a)/(1 + \bar{y}_a)$, the larger the social marginal benefits of an improved capital allocation, and thus the higher the optimal tax on capital income. The inverse relative hazard rate of the distribution of capital income, $(1 - G(y))/(yg(y))$, gives a weight to the gains of an improved allocation of capital (numerator) and the cost of distorting savings (denominator). If there are more individuals above y, $(1 - G(y))$ is larger, and the capital tax has more desirable capital reallocation effects. If the concentration of capital income at y is larger, $yg(y)$ is larger, and hence the capital tax generates larger distortions.

An alternative, but equivalent interpretation of the positive capital tax is as follows. Richer individuals obtain higher rates of return and therefore discount the future more heavily. As a result, redistribution from rich to poor is most efficient if it takes place relatively late in life. The easiest way to see this is to ignore individual behavior and consider two individuals: a poor individual with a zero marginal rate of return ($y_a = 0$) and a rich individual with a positive rate of return ($y_a > 0$). The poor individual is indifferent between receiving one additional unit of resources in the first or the second period. However, the rich would rather lose a unit of resources in the second period than in the first period. Thus, to provide the same utility gains to the poor, the government imposes smaller utility losses on the rich by taxing them in the second period rather than the first period. In other words, redistribution of second-period resources Pareto dominates redistribution of first-period resources for the same distribution of income. This is the mirror image of the scale effects generating more resources in the second period if taxes are shifted from the first to the second period. With endogenous savings decisions, taxes on capital income optimally trade off the efficiency gains of redistributing late in life with the efficiency losses from distorting savings behavior.

To the best of our knowledge, this justification for positive taxes on capital income is entirely novel. For example, Gahvari and Micheletto (2016) explicitly state that taxes on capital income are redundant if the rich earn higher returns simply because they are rich and not because they have higher ability ($y_{aa} > 0$ and $y_n = 0$ in our terminology). In other words, they find no role for taxes on capital income if return heterogeneity stems from economies of scale. This follows directly from their assumption that both taxes on labor income and capital income are levied in the same period. Proposition 3.3 shows that their result breaks down if taxes on capital income are levied later in life than taxes on labor income. In that case, the government should tax capital income if rates of return are increasing in savings.[18]

One attractive feature of the optimality condition in eq. (3.27) is that it expresses the optimal tax on capital income exclusively in terms of sufficient statistics with clear empirical counterparts. The conditional elasticity of capital income could be estimated by using exogenous variation in marginal tax rates on capital income – even though currently empirical evidence on this particular elasticity is still relatively scarce.[19] The distribution of capital income – and hence its hazard rate – can typically be obtained directly from administrative data. Moreover, there is an increasing amount of evidence on how rates of return vary with wealth (e.g., Fagereng et al., 2020). The only empirical matter on which we lack a good answer, is the extent to which heterogeneity in rates of return originates from scale economies in wealth management rather than other sources of heterogeneity such as type-dependent returns – as studied in Proposition 3.2.

We can use eq. (3.27) to determine the optimal capital tax rate at the top provided the right tail of the distribution of capital income follows a Pareto distribution. This is shown in the following Corollary.

Corollary 3.1. *Assume that capital income is convex in wealth but not directly affected by ability ($y_{aa} \geq 0$ and $y_n = 0$ for all individuals). If the right tail of the distribution of capital income follows a Pareto distribution, and if the conditional elasticity of capital income and the marginal rate of return on savings converge to the constants $\hat{e}_{y|t}$ and $\hat{y}_a$ for high levels of income, then the optimal tax rate on capital income at the top of the income distribution is constant and given by:*

$$\tau'(\hat{y}) = \frac{1}{\hat{e}_{y|z}} \frac{1}{p} \left(1 - \frac{\bar{y}_a}{\hat{y}_a}\right), \tag{3.28}$$

where a 'hat' denotes variables that refer to individuals in the top of the distribution of capital income, and $p = (1 - G(\hat{y}))/(\hat{y}g(\hat{y}))$ is the Pareto parameter of the right tail of the distribution of capital income.

Proof. Substituting $y_a = \bar{y}_a^+(y) = \hat{y}_a$, $e_{y|z} = \hat{e}_{y|z}$, and $yg(y)/(1 - G(y)) = p$ into eq. (3.27) yields eq. (3.28). □

[18]Naturally, our argument rests on the implicit assumption that taxes on labor income are levied when labor income is earned and cannot easily be deferred to later periods. Erosa and Gervais (2002) and Conesa, Kitao, and Krueger (2009) also propose positive taxes on capital income if taxes on labor income cannot be conditioned on age. However, their reasoning is very different from ours. In their model, future consumption is more complementary to leisure than current consumption. Hence, taxes on capital income are desirable to reduce distortions in labor supply (Corlett and Hague, 1953; Jacobs and Boadway, 2014; Jacobs and Rusu, 2018). This argument does not play a role in our model, because preferences are weakly separable, so that there is no reason to tax capital income to alleviate distortions in labor supply (Atkinson and Stiglitz, 1976).

[19]This holds especially when compared to the abundance of estimates of the elasticity of labor income. See Seim (2017), Zoutman (2018), and Jakobsen et al. (2018) for some recent studies on the return elasticity of wealth.

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

Eq. (3.28) allows us to make some back-of-the-envelope calculations of optimal top tax rates on capital income. Based on the data in Fagereng et al. (2020), we assume that the average rate of return is about 2.8%, whereas the (risk-adjusted) rate of return of the wealthiest decile is about 0.5 percentage points higher. We compound these returns over a period of 30 years to accommodate for the 2-period life-cycle structure of our model – see the next Section for more details. This implies a value for $\bar{y}_a/\hat{y}_a$ of $1.29/1.56 = 0.78$.[20] Provided that the right tail of the distribution of capital income is comparable to that of wealth itself, a reasonable estimate for the Pareto parameter of the U.S. distribution of capital income is $p = 1.6$ (e.g. Vermeulen, 2018).[21] Unfortunately, there are few good estimates on the elasticity of capital income. Nevertheless, a conservative estimate may be $\hat{e}_{y|z} = 0.9$, which is the baseline value in our simulations later. Taken together, these values imply an optimal top tax rate on capital income of 15 percent. This tax rate is substantial and can be much higher if lower elasticities for saving are assumed.

The general case

Propositions 3.2 and 3.3 present optimal taxes on capital income for two specific assumptions on the savings technology which are consistent with two plausible microfoundations. We can also derive an optimality condition for the general case in which capital income is a some function of savings and ability, $y^n = y(a^n, n)$, without imposing additional restrictions on the functional form of $y(\cdot)$. For this, we first write one plus the after-tax rate of return R^n as a function of capital income and ability only. To do so, we invert $y^n = y(a^n, n)$ to write $a^n = a(y^n, n)$ with $a_y = 1/y_a$ and $a_n = -y_n/y_a$. Substituting this back into R^n we obtain:

$$R^n = R(y^n, n) = 1 + (1 - \tau'(y^n))y_a(a(y^n, n), n). \tag{3.29}$$

Using this notation, we can now present an expression for the optimal tax on capital income if capital income is a general function of savings and ability.

[20] The decision whether to compound returns has little bearing on our results. Without compounding, this ratio would equal $0.028/0.033 = 0.85$.

[21] Note that the value of the Pareto parameter given here is for the distribution of capital income, whereas the value we employ in the simulations refers to the value for labor income.

Proposition 3.4. *If capital income is a general function of savings and ability $y(a^n, n)$, the optimal marginal tax rate on capital income for every level of capital income y is implied by:*

$$\begin{aligned}\left(\frac{\tau'(y)y_a}{\lambda_1/\lambda_2}\right)e_{y|z} = &\left(\frac{T'(z)}{1-T'(z)} + \frac{sy_a\tau'(y)}{\lambda_1/\lambda_2}\right)e_z\left(\frac{\xi_{y|z}}{\xi_z}\right) \\ &+ \frac{1-G(y)}{yg(y)}\left(\frac{1+\bar{y}_a^+(y)-\lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right) \\ &- \frac{1}{yg(y)}\int_y^\infty\left(\frac{\partial R(y^n,n)}{\partial n}\right)\left(\frac{\mathrm{d}y^n}{\mathrm{d}n}\right)^{-1}E_\tau^n g(y^n)\mathrm{d}y^n.\end{aligned} \tag{3.30}$$

Proof. See Appendix 3.B. □

The left-hand side and the first two terms on the right-hand side of eq. (3.30) are familiar from Propositions 3.2 and 3.3. The left-hand side represents the savings distortions associated with an increase in the tax on capital income. The first term on the right-hand side is the degree to which taxes on capital income distort labor supply less than taxes on labor income – for the same amount of income redistribution. And the second term on the right-hand side reflects the role of taxes on capital income in correcting market failures associated with scale effects. Thus, only the third term on the right-hand side of eq. (3.30) is new.

Unfortunately, the intuition behind the last term in eq. (3.30) is not straightforward. The term originates from the fact that we compare an increase in the marginal tax on labor income around z with a distributionally equivalent increase in marginal taxes on capital income. Hence, both reforms cause the same utility losses for individuals with income above z. If the discount rate is increasing with ability $\partial R(y^n, n)/\partial n > 0$, then the increase in the capital tax liability should be increasing with ability to obtain the distributional equivalent of the increase in the marginal tax on labor income. This can be achieved only if marginal taxes on capital income are raised for all levels of capital income above y for which $\partial R(y^n, n)/\partial n > 0$. These additional increases in marginal taxes yield additional distortions that are proportional to the marginal excess burden of the tax on capital income E_τ. The final term in eq. (3.30) is negative if $\partial R(y^n, n)/\partial n > 0$ and reflects this additional distortion. Naturally, the reverse holds if $\partial R(y^n, n)/\partial n < 0$, in which case the final term is positive. Hence, its sign is *a priori* ambiguous and depends on whether the marginal rate of return is increasing or decreasing in ability for a given capital income ($\partial R(y^n, n)/\partial n \gtrless 0$).

We are able to sign the term if return heterogeneity stems from a convex combination of type and scale dependency, i.e., if $y_n > 0$ and $y_{na} = 0$ as in our first microfoundation and $y_{aa} > 0$ as in our second microfoundation. To see this, take the derivative of eq. (3.29) to find:

$$\frac{\partial R(y^n, n)}{\partial n} = (1 - \tau') \left(y_{an} - \frac{y_{aa} y_n}{y_a} \right). \tag{3.31}$$

Substituting for $y_n > 0$, $y_{na} = 0$, and $y_{aa} > 0$ yields $\partial R(y^n, n)/\partial n < 0$. Intuitively, conditional on capital income, higher ability causes individuals to invest more in the closely held asset and less in the common savings technology with scale benefits. As a result, higher ability is associated with a *lower* marginal rate of return $R(y^n, n)$. Substitute this back into eq. (3.30) to find that the final term on the right-hand side is positive. Hence, the optimal tax on capital income is unambiguously positive if return heterogeneity is caused by a combination of the two microfoundations that we have discussed above.

3.6 Numerical simulation

In this Section, we calibrate our model to US data to simulate optimal non-linear taxes on labor and capital income, focussing on the case where heterogeneity in returns to capital income originates from closely-held assets.

3.6.1 Calibration

We calibrate the ability distribution and the excess returns to capital in our model to match the empirical income distribution in the US, and estimates of return heterogeneity in Norway from Fagereng et al. (2020). We use Norwegian estimates for the excess returns to capital because we are not aware of good measures for the US. The values reported by Fagereng et al. (2020) are the most comprehensive and consistent measures currently available. This combination is in line with recent work incorporating heterogeneous returns to capital in a numerical simulation, such as Guvenen et al. (2019). We calibrate the parameters of the utility function to match plausible empirical estimates for the labor supply elasticity and the intertemporal elasticity of substitution in consumption. The following subsections explain the calibration of the returns to capital, the ability distribution, behavioral elasticities for labor and capital earnings, and social preferences for income redistribution.

Timing and capital returns

We focus our simulations on the first microfoundation with closely-held assets. We parameterize the capital income function y as follows:

$$y^n \equiv ra^n + \theta(n), \tag{3.32}$$

which is equivalent to all individuals receiving a common return r and an excess return $\theta(n)/a^n$ per unit of wealth a^n. This implies $y_a = r$, $y_{aa} = 0$, $y_{an} = 0$ and $y_n \equiv \theta' > 0$.

To calibrate our two-period model, we need to map it to empirically observable yearly data. This requires assumptions regarding timing and safe interest rates. Moreover, compounding interest rates and excess returns requires special attention in our two-period framework. Each period lasts 32 years. In each year of the first period (the active period), individuals work, consume and can invest their savings at a marginal safe interest rate r^{yearly}. Hence, it is as if individuals can save in a fixed-term deposit for 32 years. Upon expiration of the deposit, individuals consume the principal and the capital income accrued over 32 years, net of taxes on capital income. Since the two periods are of equal length, and all years within a period are identical, it is as if an individual consumes in year k of the second period the savings and their returns from year k of the first period. This setup allows us to maintain the two-period structure of the theoretical model and to calibrate it using empirically observable values, such as yearly labor income and yearly returns to capital. Furthermore, the lifetime interest rate r, which is equivalent to the y_a in our theoretical model, is the yearly interest rate r^{yearly}, compounded for 32 years: $r = (1 + r^{yearly})^{32} - 1$. Thus, the government taxes income upon realization: labor income is taxed each period it is earned and capital income at the end of the 32-year holding period of each year's savings. We choose a value of yearly gross safe return of 3% per year, close to the average yearly return of the 30-year Treasury bill in the last 10 years, which was 3.28% (Federal Reserve Bank of St. Louis, 2020). By compounding this value for 32 years, we obtain the lifetime gross safe return:

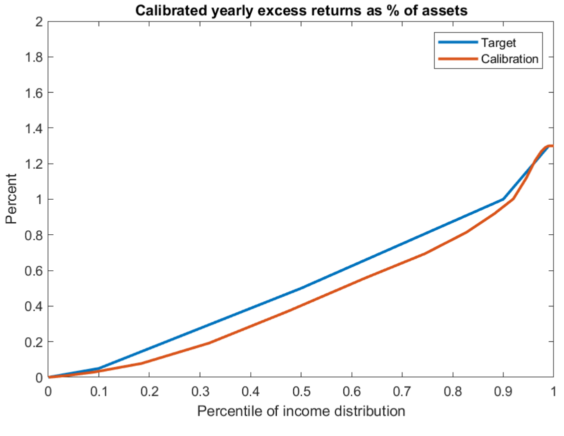
$$r = (1 + r^{yearly})^{32} - 1 = 1.57. \tag{3.33}$$

We calibrate the lifetime excess return θ to match the gradient of the yearly rate of excess returns to safe assets across the wealth distribution found, as shown by Fagereng et al. (2020) in their in Figure 3. We choose θ such that the lifetime excess <u>rate</u> of return to capital, θ/a, is the difference between the total lifetime rate of return to capital and lifetime gross safe return r in equation (3.33). The total lifetime return is calculated by compounding the yearly gross total return, which is the sum of two components: the safe rate r_{yearly} and the excess rate of return r_e^{yearly}:

$$\frac{\theta}{a} = (1 + r^{yearly} + r_e^{yearly})^{32} - 1 - r. \tag{3.34}$$

In the baseline, we choose r_e^{yearly} to match the yearly rate of excess returns in Fagereng et al. (2020) at six points of the income distribution: the 0th, 10th, 50th, 90th, 99th and 100th percentiles, and calculate the resulting θ using equation (3.34). We use a shape-preserving piecewise cubic interpolation to fill in the value of θ for the remainder of the income distribution. Fagereng et al. (2020) find that rates of return increase by about 1.3 percentage points from the poorest to the richest Norwegian individual even after controlling for differences in risk profiles. As it can be observed in Figure 3.1, our baseline calibration closely matches the gradient of the target in Fagereng et al. (2020).

Figure 3.1: Yearly excess returns in the baseline calibration



It is unclear to what extent we can generalize Fagereng et al. (2020) to other countries. For example, Bach et al. (2020) also document significant return heterogeneity in Sweden, but they argue that this can mostly be explained by differences in the risk profiles of investment portfolios. Because, as of yet, there is no consensus on the quantitative importance of type-dependent returns, we also explore the sensitivity of a lower-bound case, in which the heterogeneity in rates of return is only 20% of what is found by Fagereng et al. (2020). In that case, in our scenarios with low θ, excess returns only increase by about 0.26 percentage points across the income distribution.

Distributional assumptions

We proxy the income distribution by employing a log-normal ability distribution with an appended Pareto tail. Yearly earning ability n is log-normally distributed: $\ln n \sim N(\mu, \mathrm{sd})$, where the mean

μ and standard deviation sd of the log of ability are calibrated such that the resulting labor income distribution matches the mean and median of labor income in the US: $\$72,641$ and $\$53,657$, respectively.[22]

We append the top of the ability distribution with a Pareto tail. The Pareto parameter of the ability distribution is set to $p = 2.5$.[23,24] The starting point and the scale of the Pareto tail are chosen such that the probability density function and its first derivative are continuous. In the baseline, around 12% percent of the skill distribution is located in the Pareto tail.

Furthermore, we assume that a proportion D of individuals is disabled to guarantee that optimal marginal tax rates are always positive at the bottom. This applies to the bottom 0.15 percent of the population. Disabled individuals do not work and do not have excess returns to capital.

Utility function and behavioral elasticities

The utility function is CRRA in consumption and labor:

$$u \equiv \frac{c_1^{1-\sigma}}{1-\sigma} + \beta \frac{c_2^{1-\sigma}}{1-\sigma}, \quad \text{and} \quad v \equiv \frac{l^{\varepsilon}}{\varepsilon}, \quad \alpha, \beta, \delta, \varepsilon, \sigma > 0, \tag{3.35}$$

where $1/\sigma$ is the elasticity of intertemporal substitution (EIS), ε is the Frisch elasticity of labor supply and β captures the time preference of the individuals. Utility function (3.35) satisfies the properties we imposed on individual preferences in equation (3.5). Similar to Saez (2001) and in line with empirical estimates discussed in Flood and MaCurdy (1992), Blundell and MaCurdy (1999) and Meghir and Phillips (2010), we adopt a Frisch elasticity of labor supply ε equal to 0.22.

In our baseline scenario, the intertemporal elasticity of substitution in consumption is set to $1/\sigma = 1.5$. This is higher than some of the results reviewed by Attanasio and Weber (2010), who found estimates between 0.6 and 1, but it is a plausible value in the literature on intertemporal decision-making, which often finds estimates between 1 and 2 (Gourinchas and Parker, 2002; Mulligan, 2002; Gruber, 2006; Crossley and Low, 2011).[25] There are several reasons for choosing an EIS above 1. First, excess returns generate income effects in labor supply and saving. If the EIS is too low, income

[22] The figures represent the mean and median household income in the US in 2014 (DeNavas-Walt and Proctor, 2015; U.S. Census Bureau, 2015).

[23] We use a somewhat higher value for the Pareto parameter of labor income than conventional estimates of around 2, since the latter apply to total income (Atkinson, Piketty, and Saez, 2011). Since labor incomes are more equally distributed than capital incomes, the Pareto parameter for labor income is likely to be a bit lower.

[24] The Pareto parameter of the ability distribution is equal to the Pareto parameter of the earnings distribution if the uncompensated elasticity of earnings is zero (Saez, 2001). In particular, let the Pareto parameter of the income distribution be p^z and the uncompensated elasticity of labor supply be ε^u, then we have $p^n = p^z(1+\varepsilon^u)$. Typically, uncompensated labor elasticities are small, see Blomquist and Selin (2010) and Blundell and MaCurdy (1999).

[25] Gourinchas and Parker (2002) find an intertemporal elasticity of substitution of consumption between 0.7 and 2, while Gruber (2006) finds an estimate of 2. Some authors, such as Mulligan (2002), have estimates of $1/\sigma$ above 2. Others, such as Crossley and Low (2011), argue that the elasticity of intertemporal substitution varies across the life cycle. They report wide confidence intervals for the intertemporal elasticity of basic goods: [1.31, 2.14] for food and [0.47, 2.40] for clothing.

effects might generate a backward-bending labor supply curve, resulting in implausibly high optimal marginal tax rates. Moreover, very high marginal tax rates can lead to violations of the individual second-order conditions, so a sufficiently high EIS is required in order to match our model with empirically observed excess returns. We do allow for a lower EIS in the case where excess returns are relatively low, as income effects are also relatively small. Second, choosing a higher EIS results in more realistic responses of savings to net returns. In our baseline calibration, the resulting elasticity of capital income with respect to the net interest rate, $e^n_{y|z}$, is between 0.9 and 1.1. This is in line with the empirical estimates in Dowd, McClelland, and Muthitacharoen (2012), who find an elasticity of capital gains to capital taxes between 0.8 and 1.2. Similarly, Bruelhart et al. (2016) find a net-of-tax elasticity of capital income of 1.2 in Switzerland.

Similar to Saez (2001), we employ a linear approximation of the US tax system when calibrating the distribution of ability, the excess returns and the elasticities. The marginal tax rate on labor income is set to 35%, close to the approximation of 34% found by Heathcote, Storesletten, and Violante (2017). The marginal tax rate on capital income is set to 15%, the middle bracket of the tax on long-term capital gains in the United States. The intercept of the labor income tax is \$10,000 and the intercept of the capital tax schedule is normalized to zero. A more detailed description of the calibration algorithm can be found in Appendix 3.D.2.

Social welfare function

The government has a CARA social welfare function with inequality aversion parameter γ and a scaling parameter ρ:

$$W^n = -\int_{\underline{n}}^{\overline{n}} \frac{\rho}{\gamma} \exp\left(-U^n \frac{\gamma}{\rho}\right) f(n), \quad \gamma > 0, \quad \rho \geq 0. \tag{3.36}$$

If $\gamma \to 0$, the government is utilitarian. If $\gamma \to \infty$, the government is Rawlsian. In the baseline, we set $\gamma = 20$ to obtain, in the optimum, average income-weighted marginal tax rates of 55% for labor income and 12% for capital income. This value of the preference for redistribution makes our optimal labor tax schedule in the absence of return heterogeneity comparable to the optimal non-linear tax derived by Saez (2001) for the US, although optimal tax rates are higher than the average of marginal tax rates in the current US tax system. Saez (2001) found an average income-weighted marginal labor income tax of 59% and an asymptotic top marginal labor income tax rate of 81% with utilitarian government, allowing for income effects and assuming a compensated labor elasticity of 0.25. Our corresponding values in the case without heterogeneous returns are 55% and 74%, respectively.

3.6.2 Simulation method

Our simulations employ a structural version of our model. We express the optimal marginal tax rates in terms of model primitives such as individual choices, marginal utilities and the skill distribution, rather than in terms of sufficient statistics. Most elements of the optimal tax rules are endogenous to the non-linear tax functions that we aim to calculate. In particular, the elasticities and the income densities depend on the second-order derivatives of the tax functions, which are difficult to determine in our discrete simulation model. Solving the model formulated in terms of sufficient statistics would be computationally expensive. The optimal tax schedules in terms of model primitives are derived in Appendix 3.D.1.

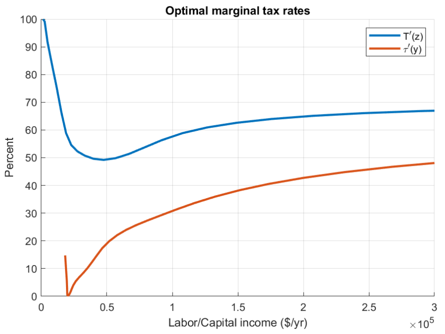
Starting from the initial tax system and an initial guess for the parameters of the skill distribution and the excess returns, we use a fixed-point algorithm to calibrate the parameters of the labor and capital income distributions such that: 1) The resulting labor income distribution has the same mean and median as the US labor income distribution and 2) The resulting yearly excess returns are in line with those found in Fagereng et al. (2020). See Appendix 3.D.2 for more details.

We extend the fixed-point algorithm developed by Mankiw, Weinzierl, and Yagan (2011) to obtain a numerical approximation of the optimal non-linear tax schedules. We start with an initial guess for the optimal tax system to calculate the second-best allocation given this initial guess. Then, the first-order conditions for the optimal taxes are computed, providing an update of the tax rates. This loop is repeated until a fixed point of the optimal tax schedule is found. For more details, see Appendix 3.D.3.

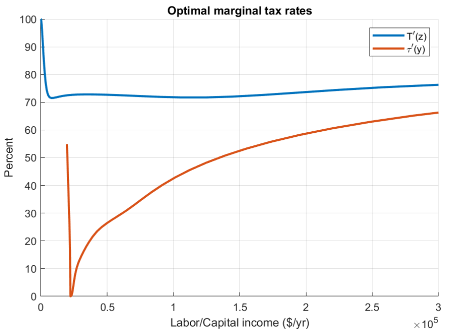
3.6.3 Optimal income taxes

We calculate optimal tax rates on labor income and capital income in Figure 3.2 for four different scenarios reflecting the substantial empirical uncertainty regarding excess returns in capital income and the elasticities of saving with respect to taxes on capital income. Doing so allows us to derive empirically plausible bounds on optimal taxes on capital income. All optimal marginal tax schedules are calculated with the baseline preference for redistribution ($\gamma = 20$).

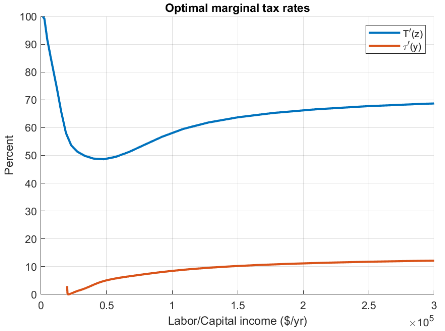
Figure 3.2: Optimal marginal income taxes on capital (red) and labor (blue)



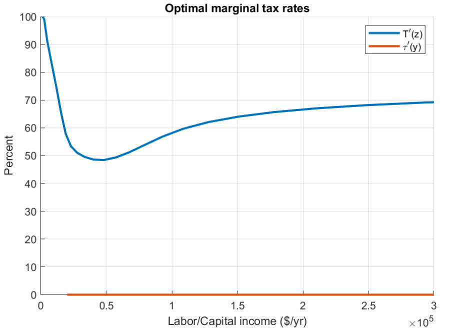
(a) Baseline: EIS = 1.5, full excess returns



(b) EIS = 0.5 (low), 20% excess returns (low)



(c) EIS = 1.5, 20% excess returns (low)



(d) Atkinson-Stiglitz: EIS = 1.5, no excess returns

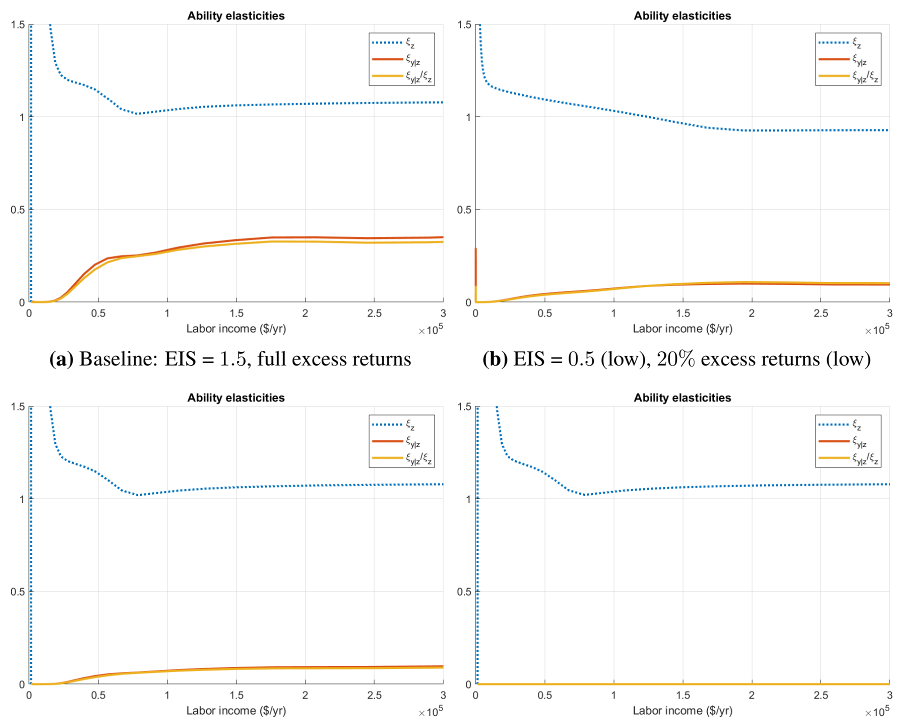
Note: the horizontal axis represents the base of each tax: yearly labor and capital income, respectively. Yearly capital income is the capital income that an individual obtains each year of the second period of their life. It consists of the total return to a deposit which has accumulated interest for 32 years.

Panel (a) covers our baseline simulation, where the elasticity of intertemporal substitution (EIS) is set at $1/\sigma = 1.5$, and excess returns are calibrated to the full magnitude found in Fagereng et al. (2020). As an alternative, panel (b) covers the case with a low elasticity of substitution: $1/\sigma = 0.5$ and excess returns are 20% of those in the baseline. In this case, saving is relatively inelastic, but excess returns are small, which avoids extremely high optimal capital tax rates due to strong income effects. This case allows us to explore the sensitivity of our results with respect to a lower EIS. Panel

(c) also decreases excess returns to 20% of those in the baseline, but maintains a high EIS. This case provides a lower bound for optimal taxes on capital income in all our simulations, where saving is relatively elastic as in the baseline, but excess returns are relatively unimportant. Panel (d) covers the Atkinson-Stiglitz case, where there are no excess returns to saving at all. This case otherwise employs the same parameters as in the baseline and should feature zero marginal capital income taxes. It also corresponds most closely to the optimal non-linear income tax of Saez (2001).

Our simulations reveal two major insights that are robust across all cases. First, the optimal marginal tax rate on capital income is positive and economically significant for most of the income distribution in all simulations with excess returns. Second, the optimal marginal tax rates on capital income increase with capital income for nearly the entire income distribution.

Figure 3.3: Ability elasticities under optimal taxation.



(a) Baseline: EIS = 1.5, full excess returns

(b) EIS = 0.5 (low), 20% excess returns (low)

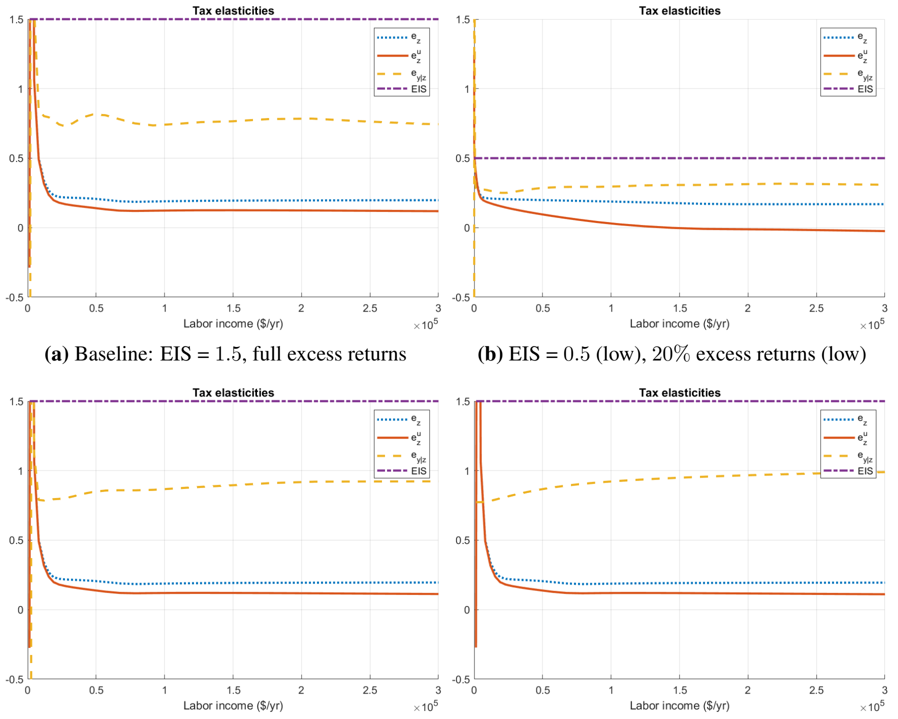
(c) EIS = 1.5, 20% excess returns (low)

(d) Atkinson-Stiglitz: EIS = 1.5, no excess returns

Optimal taxes on capital income are substantially positive in all scenarios with excess returns. The income-weighted average of the optimal marginal tax on capital income is equal to 12 percent in the baseline, about 14 percent in the case with a low EIS and low excess returns in Panel (b) in Figure 3.2 and about 2 percent in the case with a high EIS and low excess returns in Panel (c).

Optimal top tax rates on capital income converge to a positive tax rate in all simulations. The Pareto tail in the ability distribution ensures that the marginal top rate remains positive: the average income-weighted marginal rate on the top decile varies between 6 percent (high EIS, low excess returns) and 27 percent (low EIS, low excess returns), with a value of 25 percent in the baseline. The asymptotic marginal tax rate on capital income is even higher than the maximum shown in Figure 3.2. The optimal marginal tax rate converges at very high levels of capital income, i.e, after the 99.999th percentile in the baseline, whereas Figure 3.2 captures 99.86% of the distribution. The **limiting** optimal top rate on capital income varies from 14 percent for the high EIS, low excess return scenario in Panel (b) to 76 percent in the baseline scenario in Panel (a) and more than 80 percent for the low EIS, low excess returns scenario in Panel (c).

Figure 3.4: Net-of-tax elasticities under optimal taxation.



(a) Baseline: EIS = 1.5, full excess returns

(b) EIS = 0.5 (low), 20% excess returns (low)

(c) EIS = 1.5, 20% excess returns (low)

(d) Atkinson-Stiglitz: EIS = 1.5, no excess returns

Optimal marginal taxes on capital income generally increase with capital income for most parts of the income distribution. The optimal schedule may feature high marginal tax rates at the bottom, which would suggest that the optimal tax on capital income follows the U-shape of the optimal tax on labor income. However, this result is an artifact of the calibration procedure, which yields a peak in the uncompensated elasticity of labor supply ξ_z at the bottom. This is clear in Figure 3.3, which shows the elasticities that govern the optimal tax on capital income, as proven in Proposition 3.2. At the bottom of the distribution, the optimal marginal tax rate on capital income also exhibits significant variations, from 3 percent in the high EIS, low excess returns scenario in Panel (c) of Figure 3.2 to more than 50 percent in the low EIS, low excess returns scenario in Panel (b). Therefore, we do not want to claim any robustness of the high marginal tax rates at the bottom or the U-shape in optimal taxes on capital income. Moreover, the higher marginal tax rates at the bottom are in each case a very local phenomenon, compared to the gradual phasing out of the transfers in the labor income tax schedules. Indeed, our main finding is that they increase with capital income.

As the theory section has shown, the key ingredient of the optimal tax rate on capital income is the extent to which heterogeneous returns vary with ability. Figure 3.2 shows that the optimal marginal tax rate on capital income increases if excess returns in capital income become more important. This can be seen by comparing the baseline optimal taxes in Panel (a) with those in Panel (c), where excess returns are only one fifth of the baseline. This confirms our findings in section 3.5.6. Equation (3.26) shows that optimal taxes on capital income should be larger if $\xi_{y|z}/\xi_z$ is larger – ceteris paribus, where the ratio $\xi_{y|z}/\xi_z$ captures the informational content of the capital tax base relative to the labor tax base. In Figure 3.3, we explore the sensitivity of optimal capital income taxes with respect to the magnitude of return heterogeneity. The optimal marginal tax on capital income is indeed larger, the steeper the gradient of excess returns with ability, as can be seen by comparing Panels (a) and (c) in Figure 3.3: $\xi_{y|z}/\xi_z$ increases if excess returns increase.

Furthermore, a higher EIS would make capital income more responsive to taxes, thereby reducing the optimal marginal tax rates on capital income. This can be verified by comparing Panel (b) with Panel (c) in Figure 3.2, where the EIS is raised from 0.5 to 1.5 for the same excess returns. Figure 3.4 demonstrates that the conditional savings elasticity with respect to net returns $e_{y|z}$ is higher if the EIS is larger. Hence, these results confirm standard intuitions.

The robustness checks in Appendices 3.C.1 and 3.C.2 show that the optimal tax rates on both labor and capital income are lower if the government is utilitarian ($\gamma = 0$) of if the Frisch elasticity of labor supply is higher ($\varepsilon = 0.5$). Despite the lower optimal tax rates, our main conclusions that optimal marginal tax rates on capital income are positive and increasing with income remain valid.

3.7 Conclusion

A large body of empirical evidence demonstrates that people differ in the rates of return on their capital, even after controlling for risk-taking behavior. This paper analyzes optimal non-linear taxes on labor and capital income if individuals have heterogeneous capital returns. We show that optimal taxes on capital income are positive both if return differences originate from closely-held assets and if they originate from scale effects in wealth management. An empirically plausible numerical calibration of our model indicates that the optimal nonlinear tax on capital income may be significant in magnitude and increasing in income. Furthermore, the elasticity of capital income with respect to taxes is increasing in income and larger than the compensated elasticity of labor income.

Future research may extend the current paper in a number of directions. First, it would be interesting to add idiosyncratic and systematic risk and portfolio choice to analyze optimal taxes on capital income with heterogeneous returns. If portfolio choice is correlated with earnings ability, it might be optimal to distort risk-taking behavior for income redistribution. Second, this paper's two-period model structure might be extended to allow for a multiple-period life-cycle or OLG model to explore the quantitative robustness of optimal taxes on capital income in more realistic multiple-period models. Third, the model could be extended with entrepreneurial effort, tax avoidance in capital taxes and cross-border shopping. Such settings would allow for an expansion of the government instrument set to include wealth and consumption taxes.

Our paper contributes to a large and growing literature that casts doubt on the relevance of the zero capital tax results (Atkinson and Stiglitz, 1976; Chamley, 1986; Judd, 1985). Our paper provides another argument why the zero tax results offer little practical guidance for applied tax policy. Capital income should be taxed if individuals differ in the rates of return to capital, for which there is overwhelming and well-documented evidence. Hence, even if we ignore all other relevant reasons to tax capital income, the policy implications of our paper are clear: optimal taxes on capital income should be significantly positive and increasing with capital income.

3.A Proof of Lemma 3.1

We optimize the tax functions T and τ to maximize social welfare using tax perturbations. To solve this problem, we follow the Euler-Lagrange formalism, which was first applied to optimal taxation by Bohácek and Kejak (2016), Golosov, Tsyvinski, and Werquin (2014), and Lehmann et al. (2018). We start from a standard proof for the Euler-Lagrange equation, see e.g., (Arfken and Weber, 2005, ch. 17). The standard proof assumes that the arguments of the functions that are being optimized are exogenous. A complication in our case is that a change in the tax functions also affects the taxable incomes z and y. The function arguments thus depend on the functions that we are optimizing. We adapt the standard proof of the Euler-Lagrange equation to incorporate behavioral responses to tax reforms. We only prove the optimality for the tax on capital income $\tau(y)$. The proof for the tax on labor income $T(z)$ follows exactly the same steps.

We introduce the tax perturbations in Appendix 3.A.1. In Appendix 3.A.2, we study the behavioral responses to the different tax perturbations. We then prove Lemma 1 using the Euler-Lagrange approach in Appendix 3.A.3, using the results from the preceding subsections.

3.A.1 Tax reforms

Tax function $\tau(y)$ is optimal if any small perturbation of τ leaves social welfare unchanged. For any level of capital income y, we introduce a tax reform of size $\epsilon\eta(y)$. The function η is an arbitrary, non-linear, but smooth tax reform function. The parameter ϵ is infinitesimal and allows us to vary the size of the reform. We can construct any small perturbation to τ by choosing η and ϵ. The tax liability at capital income y after the tax reform becomes: $\tau(y) + \epsilon\eta(y)$. If the value of ϵ is zero, then the unperturbed tax function τ is in place. If the optimal value of ϵ is zero for every function η, then the tax schedule $\tau(y)$ is optimal.

In addition to the non-linear perturbations of size $\epsilon\eta(y)$, which we will use to characterize the optimal tax schedules, we also need to introduce scalar perturbations. The reason is that we express our characterizations of the optimal tax schedules in terms of sufficient statistics. The elasticities in Lemma 3.1 are defined as responses to scalar perturbations of size σ^T, ρ^T, σ^τ and ρ^τ. To prove the optimality of Lemma 3.1, we thus need to find relations between the behavioral responses to general perturbations of size $\epsilon\eta$ and scalar perturbations of size σ^T, ρ^T, σ^τ and ρ^τ. To do so, we reformulate the individual optimization problem taking into account both the general perturbation function and the scalar perturbation parameters. We then derive comparative statics for the individual choices.

We first rewrite the budget constraints in terms of the taxable incomes, ability, and the tax perturbations. For given ability n, there exists a one-to-one relationship between assets a and capital income $y(a, n)$. Let $a(y, n)$ be the required level of savings for a type-n individual to get capital income y. Then, the individual budget constraints are given by:

$$\mathcal{C}_1(z, y, n, \sigma^T, \rho^T) \equiv z - a(y, n) - \tilde{T}(z) - z\sigma^T - \rho^T, \tag{3.37}$$

$$\mathcal{C}_2(y, n, \sigma^\tau, \rho^\tau, \epsilon) \equiv a(y, n) + y - \tilde{\tau}(y) - \epsilon\eta(y) - y\sigma^\tau - \rho^\tau. \tag{3.38}$$

Substitute these budget constraints into the individual utility function to find the reduced-form utility function:

$$\mathcal{U}(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon) \equiv u(\mathcal{C}_1(z, y, n, \sigma^T, \rho^T), \mathcal{C}_2(y, n, \sigma^\tau, \rho^\tau, \epsilon)) - v(z/n). \tag{3.39}$$

Individuals choose labor income z and capital income y to maximize utility (3.39) subject to budget constraints (3.37) and (3.38). The first-order conditions are:

$$\mathcal{U}_z = \mathcal{U}_y = 0. \tag{3.40}$$

The functions $\mathcal{U}_z(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon)$ and $\mathcal{U}_y(z, y, n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon)$ correspond to the shift functions introduced by Jacquet, Lehmann, and Van der Linden (2013).

We denote supply functions for capital and labor income for a type-n individual for given values of the perturbation parameters as $\tilde{y}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$ and $\tilde{z}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$. We denote the corresponding indirect utility function as $\mathcal{V}(\sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon, n)$. We apply the envelope theorem to individual objective (3.39) to find the following property:

$$\mathcal{V}_\epsilon = -u_2\eta(y). \tag{3.41}$$

3.A.2 Behavioral responses to tax reforms

Suppose that there is a tax change. How do individuals change their behavior? To answer this question, suppose there is a marginal change in any parameter $\nu \in \{n, \sigma^T, \sigma^\tau, \rho^T, \rho^\tau, \epsilon\}$. Individuals adjust their behavior such that their first-order conditions remain satisfied:

$$0 = \frac{\mathrm{d}\mathcal{U}_z}{\mathrm{d}\nu} = \mathcal{U}_{zz}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} + \mathcal{U}_{zy}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} + \mathcal{U}_{z\nu}, \tag{3.42}$$

$$0 = \frac{\mathrm{d}\mathcal{U}_y}{\mathrm{d}\nu} = \mathcal{U}_{yz}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} + \mathcal{U}_{yy}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} + \mathcal{U}_{y\nu}. \tag{3.43}$$

The terms $\mathrm{d}\tilde{z}/\mathrm{d}\nu$ and $\mathrm{d}\tilde{y}/\mathrm{d}\nu$ capture the total effects of a marginal change in the parameter ν on taxable incomes. The terms $\mathrm{d}\tilde{z}/\mathrm{d}\nu$ and $\mathrm{d}\tilde{y}/\mathrm{d}\nu$ include second-round effects caused by the non-linearity of the tax schedules. If taxable incomes change due to a reform, individuals face new marginal tax rates. These changes in the marginal tax rates trigger additional behavioral responses.[26] If we write eqs. (3.42)–(3.43) in matrix notation, we can derive the following Lemma.

Lemma 3.2. *The comparative statics of a marginal change in any parameter ν on labor and capital incomes are given by:*

$$\begin{pmatrix} \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\nu} \\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\nu} \end{pmatrix} = - \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{U}_{z\nu} \\ \mathcal{U}_{y\nu} \end{pmatrix}. \tag{3.44}$$

Lemma 3.2 has the advantage that it reduces the task of finding relations between the effects of different perturbations on the tax bases, to the task of finding relations between the corresponding partial derivatives of $\mathcal{U}_z$ and $\mathcal{U}_y$. We use this result to establish the following Lemma, which shows how any capital income y, the effects of any tax perturbation $\epsilon\eta(y)$ can be decomposed into income effects and substitution effects.

Lemma 3.3. *The behavioral responses to tax perturbations $\epsilon\eta$ can be decomposed into income and substitution effects:*

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\epsilon} = \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^{\tau}}\eta(y) + \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^{\tau}} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^{\tau}}\right)\eta'(y), \tag{3.45}$$

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\epsilon} = \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^{\tau}}\eta(y) + \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^{\tau}} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^{\tau}}\right)\eta'(y). \tag{3.46}$$

[26] Jacquet, Lehmann, and Van der Linden (2013) and Jacobs and Boadway (2014) include similar second-round effects to define behavioral elasticities.

Proof. The function $\mathcal{U}$ has the following second-order partial derivatives which are evaluated at the non-reformed allocation, i.e., $\varepsilon = \rho^T = \rho^\tau = \sigma^T = \sigma^\tau = 0$:

$$\mathcal{U}_{z\rho^T} = -\frac{v' u_{11}}{n u_1}, \qquad \mathcal{U}_{y\rho^T} = \left(u_{11} - u_{12}\frac{u_1}{u_2}\right)\frac{1}{y_a}, \tag{3.47}$$

$$\mathcal{U}_{z\rho^\tau} = -\frac{v' u_{12}}{n u_1}, \qquad \mathcal{U}_{y\rho^\tau} = \left(u_{21} - u_{22}\frac{u_1}{u_2}\right)\frac{1}{y_a}, \tag{3.48}$$

$$\mathcal{U}_{z\sigma^T} = z\mathcal{U}_{z\rho^T} - u_1, \qquad \mathcal{U}_{z\sigma^\tau} = y\mathcal{U}_{z\rho^\tau}, \tag{3.49}$$

$$\mathcal{U}_{y\sigma^\tau} = y\mathcal{U}_{y\rho^\tau} - u_2, \qquad \mathcal{U}_{y\sigma^T} = z\mathcal{U}_{y\rho^T}, \tag{3.50}$$

$$\mathcal{U}_{z\epsilon} = \mathcal{U}_{z\rho^\tau}\eta(y), \qquad \mathcal{U}_{y\epsilon} = \mathcal{U}_{y\rho^\tau}\eta(y) - u_2\eta'(y). \tag{3.51}$$

$$\mathcal{U}_{zz} = -u_1 T'' + u_{11}\left(\frac{v'}{n u_1}\right)^2 - \frac{v''}{n^2}, \tag{3.52}$$

$$\mathcal{U}_{zy} = -\left(u_{11} - u_{12}\frac{u_1}{u_2}\right)\frac{1}{y_a}\frac{v'}{n u_1}, \tag{3.53}$$

$$\mathcal{U}_{yy} = \left[u_{11} - 2u_{12}\frac{u_1}{u_2} + u_{22}\left(\frac{u_1}{u_2}\right)^2 - (u_2 - u_1)\frac{y_{aa}}{y_a}\right]\frac{1}{y_a^2} - u_2\tau'', \tag{3.54}$$

$$\mathcal{U}_{zn} = \frac{v'}{n}\left[\frac{u_{11} - u_{21}}{u_1}\frac{y_n}{y_a} + \frac{1}{n}\left(1 + \frac{v'' l}{v'}\right)\right], \tag{3.55}$$

$$\mathcal{U}_{yn} = -\frac{y_n}{y_a^2}\left[(u_{11} - u_{12}) - (u_{21} - u_{22})\frac{u_1}{u_2} + (u_2 - u_1)\left(\frac{y_{an}}{y_n} - \frac{y_{aa}}{y_a}\right)\right]. \tag{3.56}$$

Verify the following relations between equations (3.48), (3.49), (3.50) and (3.51):

$$\mathcal{U}_{z\epsilon} = \mathcal{U}_{z\rho^\tau}\eta(y) + (\mathcal{U}_{z\sigma^\tau} - y\mathcal{U}_{z\rho^\tau})\eta'(y), \tag{3.57}$$

$$\mathcal{U}_{y\epsilon} = \mathcal{U}_{y\rho^\tau}\eta(y) + (\mathcal{U}_{y\sigma^\tau} - y\mathcal{U}_{y\rho^\tau})\eta'(y). \tag{3.58}$$

Substitute Lemma 3.2 for the partial derivatives of $\mathcal{U}$ to find equations (3.45) and (3.46). □

Lemma 3.3 shows that a perturbation of parameter ϵ has two effects. First, the tax liability at each capital income y increases by $\eta(y)$, which causes income effects on individual behavior. Second, the marginal tax rate at each capital income y increases by $\eta'(y)$, which causes substitution effects on individual behavior.

3.A.3 Proof of Lemma 3.1: Euler-Lagrange formalism

Consider the optimization problem of a government choosing the optimal value of ϵ, for a given reform function η. The Lagrangian for this maximization problem is given by:

$$\Lambda(\epsilon) \equiv \int_0^{\overline{n}} W(\mathcal{V}(\epsilon, n)) f(n) \mathrm{d}n + \lambda_1 \int_0^{\overline{n}} [\tilde{T}(\tilde{z}(\epsilon, n)) - g_1] f(n) \mathrm{d}n$$
$$+ \lambda_2 \int_0^{\overline{n}} [\tilde{\tau}(\tilde{y}(\epsilon, n)) + \epsilon \eta(\tilde{y}(\epsilon, n)) - g_2] f(n) \mathrm{d}n. \tag{3.59}$$

We use short-hand notations for the function arguments, ignoring the other reform parameters. We assume that this objective function is sufficiently smooth, excluding kinks and bunching. Evaluate the first-order condition for $\epsilon = 0$, using property (3.41) and Lemma 3.3:

$$\frac{\partial \Lambda(0)}{\partial \epsilon} = 0 \Leftrightarrow 0 = \int_{y^0}^{\overline{y}} \left[\frac{1}{\lambda_1/\lambda_2} - \frac{W' u_2}{\lambda_1} + T' \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} + \frac{\tau'}{\lambda_1/\lambda_2} \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} \right] \eta(y) g(y) \mathrm{d}y$$
$$+ \int_{y^0}^{\overline{y}} \left[T' \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} \right) + \frac{\tau'}{\lambda_1/\lambda_2} \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} \right) \right] \eta'(y) g(y) \mathrm{d}y. \tag{3.60}$$

To arrive at eq. (3.60), we changed the integration variables from abilities to capital incomes. To do so, we used identity $\mathrm{d}F(n) = \mathrm{d}G(y^n) \Leftrightarrow f(n)\mathrm{d}n = g(y^n)\mathrm{d}y^n$. The latter identity follows from the monotonicity of the allocation. Perform partial integration on the second line of eq. (3.60), and substitute definitions α_2 and E_τ from the main text:

$$0 = \int_{y^0}^{\overline{y}} \left[\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) + \frac{\mathrm{d}E_\tau(y) g(y)}{\mathrm{d}y} \right] \eta(y) \mathrm{d}y + E_\tau(y^0)\eta(y^0)g(y^0) - E_\tau(\overline{y})\eta(\overline{y})g(\overline{y}). \tag{3.61}$$

Eq. (3.61) must hold for every reform function η. Suppose first that the term within square brackets differs from zero on some interval. Then, choose η such that it is zero at the endpoints. The last two terms of eq. (3.61) become zero. Furthermore, let η have the same sign everywhere as the term within square brackets. With this choice of η, it follows that the value of the integral is strictly positive. However, eq. (3.61) tells us that the value of the integral must be zero for every reform function η. It follows by contradiction that the terms between the square brackets must sum to zero for every capital income y, which is an application of the fundamental theorem of the calculus of variations. From this follows the Euler-Lagrange equation for the optimal tax on capital income:

$$\forall y : \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) g(y) = -\frac{\mathrm{d}}{\mathrm{d}y} [E_\tau(y) g(y)]. \tag{3.62}$$

The integral in the first term of eq. (3.61) must thus be zero. Therefore, the last two terms of eq. (3.61) must also be zero for every reform function η. This yields the corresponding transversality conditions:

$$E_\tau(y^0)g(y^0) = 0 \text{ and } E_\tau(\overline{y})g(\overline{y}) = 0. \tag{3.63}$$

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

$$\forall z : (1-\alpha_1)h(z) = -\frac{\mathrm{d}}{\mathrm{d}z}\left[E_T(z)h(z)\right], \tag{3.64}$$

with transversality conditions:

$$E_T(0)h(0) = 0 \text{ and } E_T(\overline{z})h(\overline{z}) = 0. \tag{3.65}$$

Transversality conditions (3.63) and (3.65) form two systems of equations in the marginal tax rates at the end points. To arrive at Lemma 3.1, integrate equations (3.62) and (3.64) and use transversality conditions (3.63) and (3.65).

3.B Proof of Propositions 3.2, 3.3 and 3.4

Propositions 3.2 and 3.3 follow as special cases from Proposition 3.4. Proposition 3.4 characterizes the optimal marginal tax on capital income in terms of the optimal marginal tax on labor income. To prove Proposition 3.4, we thus need to combine the optimal tax schedules from Lemma 3.1 into a single equation. To be able to do so, we derive the Slutsky symmetry between the tax bases in Subsection 3.B.1. It relates effects of capital taxes on labor income to effects of labor taxes on capital income. Next, in Subsection 3.B.2, we rewrite the optimal capital tax in terms of first-period social welfare weights. Finally, in Subsection 3.B.3, we use the results of the preceding steps to prove Proposition 3.4.

3.B.1 Slutsky symmetry

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

Lemma 3.4. *Cross-prices responses of labor income and capital income comply to the following Slutsky symmetry:*

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} = \frac{1}{R}\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - z\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\right). \tag{3.66}$$

Proof. We first construct compensated reforms to the marginal tax rates. Denote taxable incomes in the pre-reform situation as $\tilde{z}^i \equiv \tilde{z}(0,0,0,0,0,n)$ and $\tilde{y}^i \equiv \tilde{y}(0,0,0,0,0,n)$. To determine compensated responses to an increase in marginal tax rates, we fix $\rho^T = -\sigma^T \tilde{z}^i$ and $\rho^\tau = -\sigma^\tau \tilde{y}^i$. Denote the indirect utility function in terms of the compensated perturbations as $\mathcal{V}^*(\sigma^T, \sigma^\tau, \epsilon, n) \equiv \mathcal{V}(\sigma^T, \sigma^\tau, -\sigma^T \tilde{z}^i, -\sigma^\tau \tilde{y}^i, \epsilon, n)$. We first show that these reforms are indeed compensated. Apply the envelope theorem to objective function (3.39). This yields the following properties:

$$\frac{\partial \mathcal{V}^*}{\partial \sigma^T} = \frac{\partial \mathcal{V}}{\partial \sigma^T} - \frac{\partial \mathcal{V}}{\partial \rho^T} \tilde{z}^i = -[\tilde{z}(n, \sigma^T, \sigma^\tau, -\sigma^T \tilde{z}^i, -\sigma^\tau \tilde{y}^i, 0) - \tilde{z}^i] u_1, \tag{3.67}$$

$$\frac{\partial \mathcal{V}^*}{\partial \sigma^\tau} = \frac{\partial \mathcal{V}}{\partial \sigma^\tau} - \frac{\partial \mathcal{V}}{\mathrm{d}\rho^\tau} \tilde{y}^i = -[\tilde{y}(n, \sigma^T, \sigma^\tau, -\sigma^T \tilde{z}^i, -\sigma^\tau \tilde{y}^i, 0) - \tilde{y}^i] u_2. \tag{3.68}$$

The last two expressions are zero in the situation before any reforms, when $\rho^T = \rho^\tau = \sigma^T = \sigma^\tau = 0$. We thus indeed construct compensated reforms of changes in the marginal tax rates.

Evaluate the partial derivative of eq. (3.67) with respect to σ^τ and of eq. (3.68) with respect to σ^T, both for the situation without reform:

$$\frac{\partial^2 \mathcal{V}^*}{\partial \sigma^\tau \partial \sigma^T} = -\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - \tilde{z}^i \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} \right) u_1, \tag{3.69}$$

$$\frac{\partial^2 \mathcal{V}^*}{\partial \sigma^T \partial \sigma^\tau} = -\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - \tilde{y}^i \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T} \right) u_2. \tag{3.70}$$

Young's theorem implies that the second-order derivatives of any function are symmetric. Apply this requirement to (3.69) and (3.70) to find Slutsky symmetry (3.66). □

3.B.2 Optimal tax on capital income in terms of first-period social welfare weights

To find the optimal combination of taxes on labor income and on capital income, we need to combine the optimal tax schedules in Lemma 3.1. The tax on capital income (3.24) is formulated in terms of social welfare weights in the second period, while the tax on labor income (3.23) is formulated in terms of social welfare weights in the first period. We will rewrite the optimal capital income tax in terms of first-period social welfare weights. The first step is to rewrite income effects of reforms in the second period to income effects of reforms in the first period.

The effects of changes in tax liabilities depend on the period in which they occur. If net income increases in the first period, individuals increase their savings. If net income increases in the second period instead, individuals decrease their savings. Furthermore, both responses have different effects on the marginal tax on capital income and on the marginal returns to capital since both taxes on capital income and returns to capital are non-linear. The different effects on the marginal tax rates and on the marginal returns to capital cause different second-round compensated responses. Together these differences affect the government's optimal choice between the tax bases. The following Lemma shows the relation between income effects in both periods.

Lemma 3.5. *(1) The effects of changes to unearned incomes in the first and in the second period are related as follows:*

$$R\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^{\tau}} = \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^{T}} - \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^{\tau}} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^{\tau}}\right)\frac{\partial R}{\partial y}, \tag{3.71}$$

$$R\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^{\tau}} = \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^{T}} + y_a - \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^{\tau}} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^{\tau}}\right)\frac{\partial R}{\partial y}, \tag{3.72}$$

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

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

$$\alpha_2 R = \alpha_1 - \frac{\partial R}{\partial y}E_\tau - \frac{\tau'}{\lambda_1/\lambda_2}y_a. \tag{3.73}$$

Proof. Use the individual Euler-condition (3.7) to find $1-\tau' = (u_1/u_2 - 1)/y_a$. Use this result to find:

$$\frac{\partial R}{\partial y} = -\tau'' y_a + \left(\frac{u_1}{u_2} - 1\right)\frac{y_{aa}}{y_a^2}. \tag{3.74}$$

Use eq. (3.74) to verify the following relation between equations (3.47)–(3.50) and (3.53)–(3.54):

$$\begin{pmatrix}\mathcal{U}_{z\rho^T}\\ \mathcal{U}_{y\rho^T}\end{pmatrix} = y_a\begin{pmatrix}\mathcal{U}_{zy}\\ \mathcal{U}_{yy}\end{pmatrix} + \frac{u_1}{u_2}\begin{pmatrix}\mathcal{U}_{z\rho^\tau}\\ \mathcal{U}_{y\rho^\tau}\end{pmatrix} + \frac{\partial R}{\partial y}\begin{pmatrix}\mathcal{U}_{z\sigma^\tau} - z\mathcal{U}_{z\rho^\tau}\\ \mathcal{U}_{y\sigma^\tau} - y\mathcal{U}_{y\rho^\tau}\end{pmatrix}. \tag{3.75}$$

Substitute Lemma 3.2 for the second-order derivatives of $\mathcal{U}$:

$$\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}\left[\frac{u_1}{u_2}\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} + \frac{\partial}{\partial y}\left(\frac{u_1}{u_2}\right)\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} - \begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\end{pmatrix}\right] = y_a\begin{pmatrix}\mathcal{U}_{zy}\\ \mathcal{U}_{yy}\end{pmatrix}. \tag{3.76}$$

Verify that this is equivalent to:

$$\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}\left[\frac{u_1}{u_2}\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} + \frac{\partial}{\partial y}\left(\frac{u_1}{u_2}\right)\begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\end{pmatrix} - \begin{pmatrix}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\\ \frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\end{pmatrix} - \begin{pmatrix}0\\ y_a\end{pmatrix}\right] = 0. \tag{3.77}$$

The part between square brackets must be zero. This proves the first part of the Lemma. The second part of the Lemma follows by substituting equations (3.71) and (3.72) into definition (3.21) of α_1. $\square$

We use Lemma 3.5 to rewrite the optimal capital income tax ((3.24)) in terms of first-period social welfare weights, in Lemma 3.6.

Lemma 3.6. *The optimal capital income tax can be written as follows:*

$$RE_\tau g(y^n) = \int_{y^n}^{\overline{y}} \left[1 - \alpha_1 + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2} - \left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right) E_\tau\right] g(y)\mathrm{d}y. \tag{3.78}$$

Proof. (a) The first fundamental theorem of calculus and transversality condition (3.63) yield the following expression:

$$RE_\tau g(y^n) = -\int_{y^n}^{\overline{y}} \frac{\mathrm{d}}{\mathrm{d}y}(RE_\tau g(y))\mathrm{d}y = -\int_{y^n}^{\overline{y}} \left[\frac{\mathrm{d}R}{\mathrm{d}y} E_\tau g(y) + R\frac{\mathrm{d}}{\mathrm{d}y}(E_\tau g(y))\right] \mathrm{d}y. \tag{3.79}$$

(b) Substitute eq. (3.73) into the optimal capital income tax (3.24) and use the individuals' Euler equation to find:

$$E_\tau g(y^n) = \int_{y^n}^{\overline{y}} \left[\frac{1}{\lambda_1/\lambda_2} - \frac{1}{R}\left(\alpha_1 - \frac{\partial R}{\partial y} E_\tau - \frac{\tau'}{\lambda_1/\lambda_2} y_a\right)\right] g(y)\mathrm{d}y \tag{3.80}$$

$$= \int_{y^n}^{\overline{y}} \left[\frac{1}{R}\left(1 + \frac{R}{\lambda_1/\lambda_2} - 1 - \alpha_1 + \frac{\partial R}{\partial y} E_\tau + \frac{\tau' y_a}{\lambda_1/\lambda_2}\right)\right] g(y)\mathrm{d}y \tag{3.81}$$

$$= \int_{y^n}^{\overline{y}} \frac{1}{R}\left(1 - \alpha_1 + \frac{\partial R}{\partial y} E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right) g(y)\mathrm{d}y. \tag{3.82}$$

Take the derivatives with respect to y on both sides:

$$\frac{\mathrm{d}}{\mathrm{d}y}(E_\tau g(y)) = -\frac{1}{R}\left(1 - \alpha_1 + \frac{\partial R}{\partial y} E_\tau + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right) g(y). \tag{3.83}$$

Substitute the latter into (3.79) to prove the Lemma. □

3.B.3 Proof of Proposition 3.4

We now have all the required elements to prove Proposition 3.4. We start by combining the expressions for the optimal taxes on capital and labor income. Then, we use the properties from the preceding subsections to find Proposition 3.4.

Proof. Rewrite optimal labor tax (3.23) by bringing the marginal tax on labor income to one side:

$$\begin{aligned} T'(z^n) = &- \frac{1}{h(z^n)} \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right)^{-1} \int_{z^n}^{\overline{z}} (1 - \alpha_1) f(z)\,\mathrm{d}z \\ &- \frac{\tau'}{\lambda_1/\lambda_2} \left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right)^{-1} \left(\frac{\mathrm{d}\tilde{y}^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}\tilde{y}^n}{\mathrm{d}\rho^T}\right). \end{aligned} \tag{3.84}$$

Substitute this value of T' into the optimal capital tax (3.78) and use identities $f(n) = h(z)\mathrm{d}z/\mathrm{d}n = g(y)\mathrm{d}y/\mathrm{d}n$:

$$\begin{aligned}
&\frac{\tau'(y^n)}{1+r}\left[\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right) - \left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T}\right)\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right)^{-1}\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right)\right] \\
&\quad = \left[\frac{\mathrm{d}\tilde{z}}{\mathrm{d}n}\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z^n\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right)^{-1}\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y^n\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right) - \frac{\mathrm{d}\tilde{y}}{\mathrm{d}n}\frac{u_2}{u_1}\right]\frac{1}{f(n)}\int_{y^n}^{\overline{y}}(1-\alpha_1)g(y)\mathrm{d}y \\
&\quad\quad - \frac{\mathrm{d}\tilde{y}}{\mathrm{d}n}\frac{u_2}{u_1}\frac{1}{f(n)}\int_{y^n}^{\overline{y}}\left[-\left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right)E_\tau + \frac{1+\tilde{y}_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right]g(y)\mathrm{d}y. \qquad (3.85)
\end{aligned}$$

To simplify this equation, use the following identities:

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}n} = \frac{\partial\tilde{y}}{\partial z}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}n} + \frac{\partial\tilde{y}}{\partial n}, \qquad (3.86)$$

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau} = \left(\frac{\partial\tilde{y}}{\partial\sigma^\tau} - y\frac{\partial\tilde{y}}{\mathrm{d}\rho^\tau}\right) + \frac{\partial\tilde{y}}{\partial z}\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right), \qquad (3.87)$$

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^T} - z\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^T} = \frac{\partial\tilde{y}}{\partial z}\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^T} - z\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^T}\right). \qquad (3.88)$$

We used the separability of preferences between consumption and leisure in the last equation. Substitute these three identities into eq. (3.85) and use Slutsky symmetry (3.66) to derive:

$$\begin{aligned}
&\frac{\tau' y_a^n}{1+r}\left[\frac{R}{y^n}\frac{1}{y_a^n}\left(\frac{\partial\tilde{y}}{\partial\sigma^\tau} - y\frac{\partial\tilde{y}}{\mathrm{d}\rho^\tau}\right)\right] = -\left(\frac{n}{y^n}\frac{\partial\tilde{y}}{\partial n}\right)\left(\frac{n}{z^n}\frac{\mathrm{d}\tilde{z}}{\mathrm{d}n}\right)^{-1}\frac{1}{zh(z)}\int_{z^n}^{\infty}(1-\alpha_1)h(z)\mathrm{d}z \qquad (3.89) \\
&- \frac{1}{y^n g(y^n)}\int_{y^n}^{\infty}\frac{1+\tilde{y}_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}g(y)\mathrm{d}y + \frac{1}{y^n g(y^n)}\int_{y^n}^{\infty}\left(\frac{\mathrm{d}R}{\mathrm{d}y} - \frac{\partial R}{\partial y}\right)E_\tau g(y)\mathrm{d}y.
\end{aligned}$$

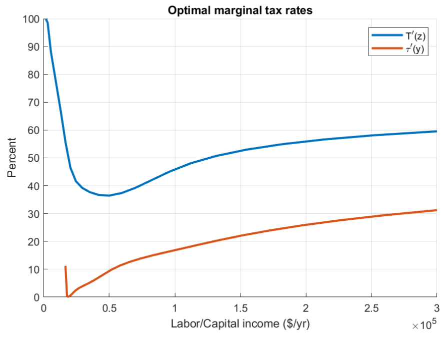
By substituting the optimal labor tax, we thus rewrote the optimal capital tax in terms of capital income elasticities conditional on labor income. Substitute the definitions for $\xi^*_{y|z}$, ξ^*_z, $e^*_{y|z}$ and e^*_z and the optimal labor tax to find Proposition 3.4. □

3.C Robustness checks

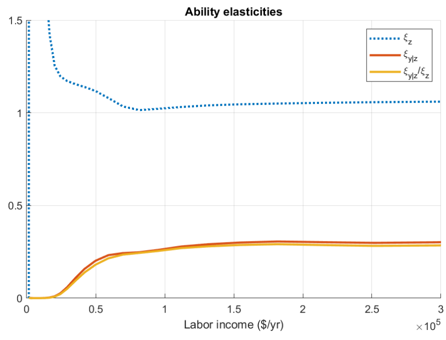
In this Section, we study how the optimal tax schedules vary with lower inequality aversion ($\gamma = 0$) and with a higher the Frisch elasticity of labor supply ($\varepsilon = 0.5$). We find that optimal marginal tax rates decrease if we decrease the inequality aversion, and if we increase the Frisch elasticity. Our main conclusions remain valid: optimal marginal taxes on capital income are substantial and they increase with income.

3.C.1 Overview of results with utilitarian preferences ($\gamma = 0$)

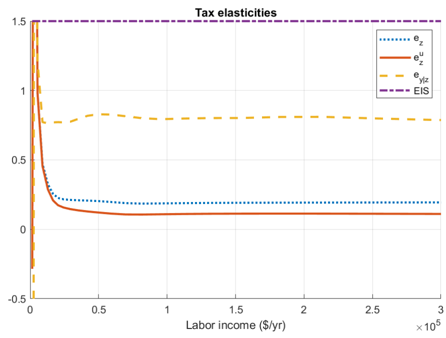
Figure 3.5: Results with utilitarian government ($\gamma = 0$) in the baseline (high EIS, full excess returns)



(a) Optimal marginal capital and labor income taxes



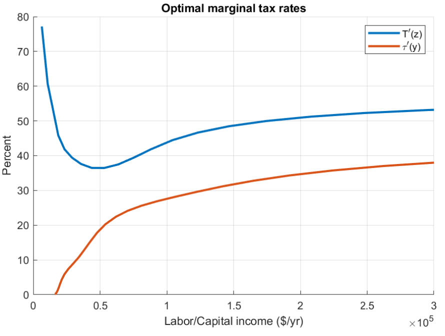
(b) Ablity elasticities



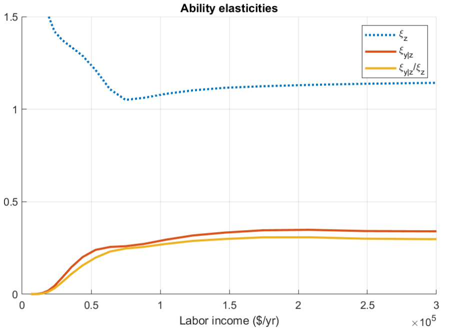
(c) Net-of-tax elasticites

3.C.2 Overview of results with high Frisch elasticity ($\varepsilon = 0.5$)

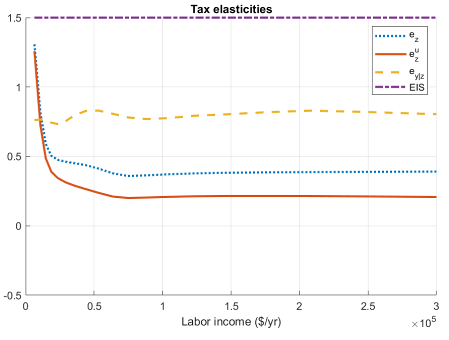
Figure 3.6: Results with high Frisch elasticity ($\varepsilon = 0.5$) in the baseline (high EIS, full excess returns)



(a) Optimal marginal capital and labor income taxes



(b) Ablity elasticities



(c) Net-of-tax elasticites

3.D Algorithms for the simulations

The numerical simulations consist of two parts: the calibration of the model primitives and the calculation of the optimal non-linear tax schedules. We explain the theoretical preliminaries for our structural approach in Subsection 3.D.1. We then explain the algorithm for the calibration in Subsection 3.D.2, and the algorithm for the actual simulations in Subsection 3.D.3.

3.D.1 Theoretical preliminaries

So far, we have written the optimal tax schedules in terms of sufficient statistics. However, this section develops a structural model to numerically simulate the optimal tax schedules. The reason for doing so is that the optimal tax expressions in sufficient statistics depend on the second-order derivatives of the tax functions T and τ.[27] Approximating the second-order derivatives of the tax functions is computationally expensive. Hence, in the numerical simulations we express the optimal tax schedules in terms of the model primitives: the utility function, the social welfare function, the individual abilities and their distribution, and excess returns to saving.

We start in Subsections 3.D.1 and 3.D.1 by formally introducing a mass of disabled individuals at the bottom to ensure no-zero marginal tax rates at the bottom. Next, in Subsection 3.D.1, we study the relation between the multipliers on the government budget constraint λ_1 and λ_2. Then, in Subsection 3.D.1, we reformulate the optimal tax schedules in terms of model primitives, for a given social marginal value of income redistribution. In Subsection 3.D.1 we reformulate the social marginal value of income redistribution in terms of model primitives. We find the optimal value for the first-period government budget multiplier λ_1 in Subsection 3.D.1.

Disabled individuals

A fraction D of the population is disabled. We denote variables for a disabled individual using a superscript d. Disabled individuals have zero earning ability ($n = 0$) and do not earn any labor income: $z^d = 0$. In the first period, disabled individuals receive a transfer $-T(0)$, which they consume immediately or save for consumption in the second period. The level of savings of the disabled is equal to the level of savings of the least able individual if the tax system is optimized: $a^d = a^0$. To keep the simulations tractable, the savings of the disabled are exogenous: they do not respond to tax perturbations. Neither the disabled, nor the least able individual earn any excess

[27] Lemma 3.2 demonstrates that the sufficient statistics depend on the Hessian of the reduced-form utility function $\mathcal{U}$, while the proof of Lemma 3.3 shows that the elements of this Hessian depend on the second-order derivatives of the tax functions T and τ.

returns: $y^d = y^0 = ra^0$. The disabled pay the same tax as the workers with the same capital income $-\tau(y^d)$.

The budget constraints for the disabled individuals are:

$$\mathcal{C}_1^d(y^d, \rho^T) = -\tilde{T}(0) - \rho^T - \frac{y^d}{r}, \tag{3.90}$$

$$\mathcal{C}_2^d(y^d, \sigma^\tau, \rho^\tau, \epsilon) = \frac{y^d}{r} + y^d - \tilde{\tau}(y^d) - \epsilon\eta(y^d) - y^d\sigma^\tau - \rho^\tau. \tag{3.91}$$

We denote the indirect utility function for the disabled as $\mathcal{V}^d(\sigma^\tau, \rho^T, \rho^\tau, \epsilon)$. Use the envelope theorem to find the following properties:

$$\mathcal{V}_\epsilon^d = -u_2^d\eta(y^d),\ \mathcal{V}_{\rho^\tau}^d = -u_2^d. \tag{3.92}$$

Optimal tax schedules with disabled individuals

The new Lagrangian for the government optimization problem, taking into account the mass of individuals at the bottom, is given by:

$$\begin{aligned}\Lambda(\epsilon) \equiv & DW(\mathcal{V}^d(\epsilon)) + \int_0^{\overline{n}} W(\mathcal{V}(\epsilon, n))f(n)\mathrm{d}n + \lambda_1[D\tilde{T}(0) - g_1] + \lambda_1 \int_0^{\overline{n}} \tilde{T}(\tilde{z}(\epsilon, n))f(n)\mathrm{d}n \\ & + \lambda_2 D[\tilde{\tau}(y^d) + \epsilon\eta(y^d)] - \lambda_2 g_2 + \lambda_2 \int_0^{\overline{n}} [\tilde{\tau}(\tilde{y}(\epsilon, n)) + \epsilon\eta(\tilde{y}(\epsilon, n))]f(n)\mathrm{d}n.\end{aligned} \tag{3.93}$$

We again assume that this objective function is sufficiently smooth, excluding kinks and bunching. Evaluate the first-order condition at $\epsilon = 0$, using properties (3.41) and (3.92) and Lemma 3.3:

$$\begin{aligned}\frac{\partial\Lambda(0)}{\partial\epsilon} = 0 \Leftrightarrow 0 = & \left(\frac{1}{\lambda_1/\lambda_2} - \frac{W'u_2^d}{\lambda_1}\right)\eta(y^0)D \\ & + \int_{y^0}^{\overline{y}} \left[\frac{1}{\lambda_1/\lambda_2} - \frac{W'u_2}{\lambda_1} + T'\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau} + \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right]\eta(y)g(y)\mathrm{d}y \\ & + \int_{y^0}^{\overline{y}} \left[T'\left(\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{z}}{\mathrm{d}\rho^\tau}\right) + \frac{\tau'}{\lambda_1/\lambda_2}\left(\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\sigma^\tau} - y\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\rho^\tau}\right)\right]\eta'(y)g(y)\mathrm{d}y.\end{aligned} \tag{3.94}$$

Perform partial integration on the second line of eq. (3.94) and substitute definitions α_2 and E_τ from the main text:

$$\begin{aligned}0 = & \int_{y^0}^{\overline{y}} \left[\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2\right)g(y) + \frac{\mathrm{d}E_\tau(y)g(y)}{\mathrm{d}y}\right]\eta(y)\mathrm{d}y \\ & + \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^d\right)\eta(y^0)D + E_\tau(y^0)\eta(y^0)g(y^0) - E_\tau(\overline{y})\eta(\overline{y})g(\overline{y}).\end{aligned} \tag{3.95}$$

Applying the fundamental theorem of the calculus of variations in the same way as we did in Subsection 3.A.3, we find that the Euler-Lagrange Equation (3.62) for the optimal tax on capital income is not directly affected by the introduction of a mass of individuals at the bottom. Furthermore, the transversality condition (3.63) at the top remains unaltered. At the bottom though, we find the following transversality condition:

$$E_\tau(y^0)g(y^0) = -\left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2^d\right) D. \tag{3.96}$$

Entirely analogous derivations show that the Euler-Lagrange equation (3.64) for the tax on labor income and the transversality condition (3.64) at the top remain unaltered with the addition of the disabled individuals. At the bottom, we find the following transversality condition for the tax on labor income:

$$E_T(0)h(0) = -(1 - \alpha_1^d)D. \tag{3.97}$$

Given that the Euler-Lagrange equations and the transversality conditions at the top are not directly affected by the presence of a mass of individuals at the bottom, the optimal tax schedules in Lemma 3.1 remain valid. Transversality conditions (3.96) and (3.97) form a new system of equations in the marginal tax rates at the bottom. The optimal marginal tax rates at the bottom are now positive, as we demonstrate in subsection 3.D.1.

Relation between the government budget multipliers

For the simulations we need explicit expressions for the multipliers on the government budget constraints λ_1 and λ_2. In this section, we study the relation between the two multipliers. Substitute Lemmas 3.5 and 3.6 into transversality condition (3.96) for the lowest ability, use our assumption that the disabled do not respond to tax perturbations, and use $\mathrm{d}R/\mathrm{d}y = \partial R/\partial y$ in the case with closely-held assets:

$$\int_0^{\overline{n}} \left(1 - \alpha_1 + \frac{1 + y_a - \lambda_1/\lambda_2}{\lambda_1/\lambda_2}\right) f(n)\mathrm{d}n = \left(\alpha_1^d - \frac{1 + y_a^d}{\lambda_1/\lambda_2}\right) D. \tag{3.98}$$

Substitute transversality condition (3.97) for the lowest ability:

$$\int_0^{\overline{n}} \left(\frac{1 + y_a}{\lambda_1/\lambda_2} - 1\right) f(n)\mathrm{d}n = \left(1 - \frac{1 + y_a^d}{\lambda_1/\lambda_2}\right) D. \tag{3.99}$$

In the case with closely-held assets, our assumptions imply that all individuals earn the same marginal return to capital y_a. It follows that in the optimum the multipliers on the government budget constraint are related as follows:

$$\lambda_1/\lambda_2 = 1 + y_a. \tag{3.100}$$

In the optimum, the marginal rate of substitution between the two periods for the government should thus be equal to the private marginal rate of transformation. Substituting eq. (3.100) into Lemma 3.6, we then find that in the case of a closely-held asset, a simple relation exists between the social welfare weights in both periods:

$$\int_n^{\overline{n}} \left(\frac{1}{\lambda_1/\lambda_2} - \alpha_2 \right) f(\hat{n}) \mathrm{d}\hat{n} = \frac{u_2}{u_1} \int_n^{\overline{n}} (1 - \alpha_1) f(\hat{n}) \mathrm{d}\hat{n}. \tag{3.101}$$

Tax schedules for given benefits of redistribution

We take two steps to rewrite the optimal marginal tax schedules in terms of model primitives. In this Subsection, we take the net social marginal benefit of a higher marginal tax rate at n, $\int_n^{\overline{n}}(1 - \alpha_1)f(\hat{n})\mathrm{d}\hat{n}$, as given. In the next Subsection we also derive the net social marginal benefits of a higher tax rate in terms of model primitives.

Rewrite Lemma 3.1 in matrix notation, taking into account Lemma 3.6, using the assumption $y_{aa} = y_{an} = 0$, and using eq. (3.100):

$$\begin{pmatrix} \frac{\mathrm{d}z^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} & \frac{\mathrm{d}z^n}{\mathrm{d}\sigma^\tau} - y^n \frac{\mathrm{d}z^n}{\mathrm{d}\rho^\tau} \\ \frac{\mathrm{d}y^n}{\mathrm{d}\sigma^T} - z^n \frac{\mathrm{d}y^n}{\mathrm{d}\rho^T} & \frac{\mathrm{d}y^n}{\mathrm{d}\sigma^\tau} - y^n \frac{\mathrm{d}y^n}{\mathrm{d}\rho^\tau} \end{pmatrix}^{tr} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = - \begin{pmatrix} 1 & 0 \\ 0 & \frac{u_2}{u_1} \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}z}{\mathrm{d}n} \\ \frac{\mathrm{d}y}{\mathrm{d}n} \end{pmatrix} \frac{\int_n^{\overline{n}}(1 - \alpha_1)f(\hat{n})\mathrm{d}\hat{n}}{f(n)}, \tag{3.102}$$

where superscript tr stands for matrix transpose. Substitute Lemma 3.2:

$$\begin{aligned} &\begin{pmatrix} \mathcal{U}_{z\sigma^T} - z^n \mathcal{U}_{z\rho^T} & \mathcal{U}_{y\sigma^T} - z^n \mathcal{U}_{y\rho^T} \\ \mathcal{U}_{z\sigma^\tau} - y^n \mathcal{U}_{z\rho^\tau} & \mathcal{U}_{y\sigma^\tau} - y^n \mathcal{U}_{y\rho^\tau} \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} \\ &\qquad = - \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{R} \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{U}_{zn} \\ \mathcal{U}_{yn} \end{pmatrix} \frac{\int_n^{\overline{n}}(1 - \alpha_1)f(\hat{n})\mathrm{d}\hat{n}}{f(n)}. \end{aligned} \tag{3.103}$$

Substitute equations (3.49)–(3.50):

$$\begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} \begin{pmatrix} \mathcal{U}_{zz} & \mathcal{U}_{zy} \\ \mathcal{U}_{yz} & \mathcal{U}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} \frac{\int_n^{\overline{n}}(1 - \alpha_1)f(\hat{n})\mathrm{d}\hat{n}}{f(n)}. \tag{3.104}$$

Eliminate the two left-most matrices on both sides to find the following characterization of the optimal tax schedules:

$$\begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = \begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} \frac{\int_n^{\overline{n}} (1-\alpha_1) f(\hat{n}) \mathrm{d}\hat{n}}{f(n)}. \tag{3.105}$$

Equation (3.105) is very similar to the optimal tax expressions in Mirrlees (1976). The main difference is that Mirrlees (1976) only accounts for the dependence of the marginal rates of substitution on ability, capturing e.g., varying degrees of complementarity between goods and leisure (Corlett and Hague, 1953), while equation (3.105) also accounts for the direct dependence of the individual budget constraints (3.37)–(3.38) on ability.

We can now derive the optimal marginal tax rates for the individuals with the lowest ability. Substitute the transversality condition for the lowest ability from eq. (3.97) into eq. (3.105) to find :

$$\begin{pmatrix} T'(0) \\ \frac{\tau'(y^0)}{1+y_a} \end{pmatrix} = -\begin{pmatrix} \frac{\mathcal{U}_{zn}}{u_1} \\ \frac{\mathcal{U}_{yn}}{u_1} \end{pmatrix} (1-\alpha_1^d) \frac{D}{f(0)}. \tag{3.106}$$

Marginal tax rates at the bottom are thus positive if single crossing conditions $\mathcal{U}_{zn}, \mathcal{U}_{yn} > 0$ are met, and social welfare weights α_1^d for the disabled are larger than one. Substitute equation (3.100) into equation (3.98) to find $1 = \alpha_1^d D + \int_0^{\overline{n}} \alpha_1 f(n) \mathrm{d}n$. If we assume that the welfare weights of the disabled are higher than the welfare weights of the able, then it follows that $\alpha_1^d > 1$, and thus marginal tax rates at the bottom are positive. Positive marginal tax rates at the bottom enable redistribution from the able to the disabled. If the proportion of disabled individuals equals zero, $D = 0$, the marginal tax rate at the bottom is zero.

Social marginal benefit of income redistribution

Equation (3.105) is not completely written in terms of model primitives. The social marginal value of a higher tax rate still contains sufficient statistics:

$$\int_n^{\overline{n}} (1-\alpha_1) f(\tilde{n}) \mathrm{d}\tilde{n} = \int_n^{\overline{n}} \left(1 - \frac{W' u_1}{\lambda_1} + T' \frac{\mathrm{d}z^{\hat{n}}}{\mathrm{d}\rho^T} + \frac{\tau'}{\lambda_1/\lambda_2} \frac{\mathrm{d}z^{\hat{n}}}{\mathrm{d}\rho^T}\right) f(\hat{n}) \mathrm{d}\hat{n}. \tag{3.107}$$

To find an alternative formulation for eq. (3.107), we rewrite it as a differential equation. Introduce the following composite function:

$$\Phi(n) \equiv -\frac{1}{u_1} \int_n^{\overline{n}} (1-\alpha_1) f(\hat{n}) \mathrm{d}\hat{n}, \tag{3.108}$$

with derivative:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}n} = -\frac{\mathrm{d}u_1}{\mathrm{d}n}\frac{\Phi}{u_1} + \frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1} + T'\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} + \frac{\tau'}{\lambda_1/\lambda_2}\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T}\right) f(n). \tag{3.109}$$

Substitute eq. (3.105) to rewrite the terms with the income effects, and use definition (3.39) to expand $\mathrm{d}u_1/\mathrm{d}n$:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}n} = \left(\mathcal{U}_{z\rho^T}\frac{\mathrm{d}z^n}{\mathrm{d}n} + \mathcal{U}_{y\rho^T}\frac{\mathrm{d}y^n}{\mathrm{d}n} + u_{11}\frac{y_n}{r}\right)\frac{\Phi}{u_1} + \frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1}\right) f(n) - \frac{\Phi}{u_1}\left(\mathcal{U}_{zn}\frac{\mathrm{d}z^n}{\mathrm{d}\rho^T} + \mathcal{U}_{yn}\frac{\mathrm{d}y^n}{\mathrm{d}\rho^T}\right). \tag{3.110}$$

Eq. (3.110) is a differential equation with Φ as unknown function. To further simplify this differential equation, we first formulate a Lemma that relates the effects of perturbations of any two parameters μ and ν on the tax bases.

Lemma 3.7. *For any two perturbation parameters μ and ν, the effects on the taxable incomes are related as follows:*

$$\mathcal{U}_{z\mu}\frac{\mathrm{d}z^n}{\mathrm{d}\nu} + \mathcal{U}_{y\mu}\frac{\mathrm{d}y^n}{\mathrm{d}\nu} = \mathcal{U}_{z\nu}\frac{\mathrm{d}z^n}{\mathrm{d}\mu} + \mathcal{U}_{y\nu}\frac{\mathrm{d}y^n}{\mathrm{d}\mu}.$$

Proof. Use Lemma 3.2 to find the following relation:

$$\begin{aligned}
\mathcal{U}_{z\mu}\frac{\mathrm{d}z^n}{\mathrm{d}\nu} + \mathcal{U}_{y\mu}\frac{\mathrm{d}y^n}{\mathrm{d}\nu} &= -\begin{pmatrix}\mathcal{U}_{z\mu} & \mathcal{U}_{y\mu}\end{pmatrix}\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz} & \mathcal{U}_{yy}\end{pmatrix}^{-1}\begin{pmatrix}\mathcal{U}_{z\nu}\\ \mathcal{U}_{y\nu}\end{pmatrix}\\
= \left[-\begin{pmatrix}\mathcal{U}_{zz} & \mathcal{U}_{zy}\\ \mathcal{U}_{yz}\mathcal{U}_{yy} & \end{pmatrix}^{-1}\begin{pmatrix}\mathcal{U}_{z\mu}\\ \mathcal{U}_{y\mu}\end{pmatrix}\right]^{tr}&\begin{pmatrix}\mathcal{U}_{z\nu}\\ \mathcal{U}_{y\nu}\end{pmatrix} = \mathcal{U}_{z\nu}\frac{\mathrm{d}z^n}{\mathrm{d}\mu} + \mathcal{U}_{y\nu}\frac{\mathrm{d}y^n}{\mathrm{d}\mu}.
\end{aligned} \tag{3.111}$$

□

Lemma 3.7 allows us to simplify differential equation (3.110):

$$\frac{\mathrm{d}\Phi(n)}{\mathrm{d}n} = -\frac{u_{11}}{u_1}\frac{y_n}{r}\Phi(n) + \frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1}\right) f(n). \tag{3.112}$$

This differential equation has a standard form, with solution:

$$\int_{n}^{\overline{n}}(1-\alpha_1)f(n^*)\mathrm{d}n^* = u_1\int_{n}^{\overline{n}}\frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1}\right)\exp\left[\int_{n}^{n^*}\frac{u_{11}}{u_1}\frac{y_{n^{**}}}{r}\mathrm{d}n^{**}\right] f(n^*)\,\mathrm{d}n^*. \tag{3.113}$$

We have thus rewritten the social marginal benefits of a higher marginal tax rate in terms of economic fundamentals. Substituting eq. (3.113) into eq. (3.105), substituting Eqs. (3.55)–(3.56) and using the assumption $u_{12} = 0$, we thus find the optimal tax schedules in terms of model primitives:

$$\begin{pmatrix} T' \\ \frac{\tau'}{1+y_a} \end{pmatrix} = \begin{pmatrix} \frac{v'}{n}\left(\frac{u_{11}}{u_1}\frac{y_n}{r} + \frac{1}{n}\left(1 + \frac{v''l}{v'}\right)\right) \\ -\frac{y_n}{r^2}\left(u_{11} + \frac{u_1}{u_2}u_{22}\right) \end{pmatrix} \frac{\int_n^{\overline{n}} \frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1}\right)\exp\left[\int_n^{n^*} \frac{u_{11}}{u_1}\frac{y_{n^{**}}}{r}\mathrm{d}n^{**}\right] f\left(n^*\right)\mathrm{d}n^*}{f(n)}. \tag{3.114}$$

Value of the first-period multiplier on the government budget constraint

Equation (3.114) characterizes the optimal tax schedules almost entirely in terms of model primitives. The only remaining variable to rewrite is the first-period multiplier on the government budget constraint. Substitute the first-period benefit of redistribution eq. (3.113) into transversality condition eq. (3.97) for the least able individual:

$$\int_0^{\overline{n}} \frac{1}{u_1}\left(1 - \frac{W'u_1}{\lambda_1}\right)\exp\left[\int_0^n \frac{u_{11}}{u_1}\frac{y_{\hat{n}}}{r}\mathrm{d}\hat{n}\right] f\left(n\right)\mathrm{d}n = -\frac{1}{u_1^0}\left(1 - \frac{W'^d u_1^d}{\lambda_1}\right)D. \tag{3.115}$$

Solve this for λ_1:

$$\lambda_1 = \frac{\int_0^{\overline{n}} W'\exp\left[\int_0^n \frac{u_{11}}{u_1}\frac{y_{\hat{n}}}{r}\mathrm{d}\hat{n}\right] f\left(n\right)\mathrm{d}n + W'^d\frac{u_1^d}{u_1^0}D.}{\int_0^{\overline{n}} \frac{1}{u_1}\exp\left[\int_0^n \frac{u_{11}}{u_1}\frac{y_{\hat{n}}}{r}\mathrm{d}\hat{n}\right] f\left(n\right)\mathrm{d}n + \frac{1}{u_1^0}} \tag{3.116}$$

The larger is the fraction of disabled individuals D, the larger will be the multiplier λ_1, and the lower will be welfare weights α_1 of the able individuals. Therefore, the larger is the proportion of disabled individuals, the higher will be the marginal tax rates for the rest of the population.

3.D.2 Calibration

We choose the parameters σ, β and ε for utility function (3.35) so that the model implies realistic values for the elasticity of intertemporal substitution, the rate of time preference and the Frisch elasticity of labor supply. Moreover, we use a fixed-point algorithm to calibrate the parameters of the skill distribution and the excess returns to capital such that the implied distributions of labor income and of the excess returns match their empirical counterparts when a linear approximation of the US tax system is in place. Table 3.1 lists our choices for the relevant parameters. Our targets for the income distribution and the excess returns are discussed in Section 3.6.1 in of the main text.

The skill distribution consists of two parts. The bulk of the population resides in the first part, where the skills follow a lognormal distribution with parameters μ and sd. The upper tail of the skill distribution resides in the second part, where the skills follow a Pareto distribution with Pareto parameter p^n. The start and the scaling parameter of the Pareto distribution are chosen such that the entire skill distribution $f(n)$ is continuous and has a continuous derivative everywhere. The

Table 3.1: Main parameters used in the baseline calibration

Symbol	Description	Value
r^{yearly}	Yearly risk-free return	0.03
ε	Frisch elasticity of labor supply	0.22
EIS	Elasticity of intertemporal substitution	1.50
β	Rate of time preference	$\beta = 1/(1+r)$
τ_e'	Empirical marginal tax on capital	0.15
T_e'	Empirical marginal tax on labor	0.35
$-T_e(0)$	Empirical lump-sum transfer	$10,000$ USD
γ	Inequality aversion of the government	20

Pareto parameter p^n of the skill distribution is chosen to match Pareto parameter p^z of the US income distribution, according to the formula $p^n = p^z(1+\varepsilon^u)$, as explained in Subsection 3.6.1 in the main text.

We use a nested fixed point algorithm to find the parameters μ and sd of the skill distribution, and the excess returns $\theta(n)$ for each skill level. In the outer loop, we try different values of μ and sd, such that the resulting moments of the labor income distribution match their targets. In the inner loop, we seek the excess returns $\theta(n)$ to match the empirical values observed by Fagereng et al. (2020), updating individual behavior as we try different values.

The entire procedure is summarized in Algorithm 1. We start from an initial guess for the distribution parameters μ and sd of the log–normal distribution. We append the Pareto tail, and we initialize excess returns $\theta(n)$ to zero for each individual. Then, given our linear approximation of the US tax system, we simulate the behavior of each individual, yielding values for their incomes z^n and their savings a^n. Given this information, we assign values to $\theta(n)$ at the percentiles given in Subsection 3.6.1 to match the excess returns $\theta(n)/a^n$ observed by Fagereng et al. (2020), and we use a smooth interpolation to assign excess returns to the intermediate skill levels. Given the new values for the excess returns $\theta(n)$, still for the same guess of distribution parameters μ and sd, we update individual behavior and the resulting excess returns until the excess returns converge to stable values. We repeat the whole procedure, updating distribution parameters μ and sd, appending the Pareto tail and calibrating the excess returns, until the moments of the resulting labor income distribution match their target values.

3.D.3 Overview of the simulation algorithm

The steps for our simulation algorithm are detailed in Algorithm 2. Starting from an initial guess for the tax system, we simulate the behavior of the individuals. Given the individual choices, we

Algorithm 1 Calibration algorithm

0. Start from initial guess for μ and sd.
1. Append Pareto tail using parameter $p^n = p^z(1 + \varepsilon^u)$.
2. Find $\theta(n)$ to match target values for the excess rates of return:
   (0) set $\theta(n) = 0$
   (a) solve the individuals' problem taking $\theta(n)$ as given, find labor incomes z and savings a
   (b) set $\theta(n)$ at chosen percentiles so excess returns $\theta(n)/a$ match their targets
   (c) use smooth interpolation to impute intermediate values of $\theta(n)$
   (d) repeat steps 2a-2c until behavior stabilizes
3. Update μ and sd, repeat steps 1-2 until the resulting labor income distribution matches its targets.

use transversality condition (3.116) to find the value of the Lagrange multiplier for the government's first-period budget constraint. Substituting the individual choices and the Lagrange multiplier for the government's first-period budget constraint into the government's first-order conditions (3.114), we find updated values for the marginal tax rates. Given the new values for the marginal tax rates, we look for new values for the tax intercepts, such that the government's budget constraints (3.16)–(3.17) are satisfied in both periods. We repeat this entire procedure, starting each iteration from the outcome of the last, until the simulations converge to the optimal schedules for the taxes on capital income and labor income.

Algorithm 2 Simulation algorithm

Let subscripts (i) refer to values in the i-th iteration of the fixed-point algorithm. The simulation algorithm then follows the following steps:

0. Start from an initial guess for the tax schedules $T_{(0)}$ and $\tau_{(0)}$.
1. Iteration i: simulate individual choices $\{c^n_{1(i)}, c^n_{2(i)}, z^n_{(i)}, y^n_{(i)}\}^{\bar{n}}_{\underline{n}}$, given the tax schedules $T_{(i)}$ and $\tau_{(i)}$.
2. Calculate Lagrange multiplier $\lambda_{1(i)}$ of the government's first-period budget constraint by substituting individual choices $\{c^n_{1(i)}, c^n_{2(i)}, z^n_{(i)}, y^n_{(i)}\}^{\bar{n}}_{\underline{n}}$ into transversality condition (3.116).
3. Substitute individual choices $\{c^n_{1(i)}, c^n_{2(i)}, z^n_{(i)}, y^n_{(i)}\}^{\bar{n}}_{\underline{n}}$ and Lagrange multiplier $\lambda_{1(i)}$ into optimal tax condition (3.114) to find updated marginal tax rates. To avoid large jumps from one iteration to the next, set the marginal tax rates $T'_{(i+1)}$ and $\tau'_{(i+1)}$ for the next iteration to a weighted average of the current and the newly calculated marginal tax rates.
4. To find the intercepts $T_{(i+1)}(0)$ and $\tau_{(i+1)}(0)$, repeat the following steps:
   (0) initialize $T_{(i+1)}(0) = T_{(i)}(0)$ and $\tau_{(i+1)}(0) = \tau_{(i)}(0)$
   (a) simulate individual behavior using the current guess for $T_{(i+1)}$ and $\tau_{(i+1)}$
   (b) given the individual choices calculated in step 4a, calculate tax revenues for the current guess for $T_{(i+1)}$ and $\tau_{(i+1)}$
   (c) repeat steps 4a-4b to find values of $T_{(i+1)}(0)$ and $\tau_{(i+1)}(0)$ that satisfies government budget constraints (3.16)–(3.17)
5. Repeat steps 1-4 until the tax schedules converge.

Chapter 4

You Sponsor Mine, I Procure Yours: Pharmaceutical Sponsorships and Procurement in Public Hospitals

4.1 Introduction

Do public servants use public funds to deliver public goods efficiently? Or is money lost due to graft, bribery, preferential treatment and influence peddling? Such questions are particularly relevant for the healthcare sector, where public funds cover a large share of total spending[1] and are rapidly expanding in low- and middle- income countries (Cotlear et al., 2015).

Empirical evidence from infrastructure spending suggests corrupt practices are associated with a lower level of public goods provision (Castro, Guccio, and Rizzo, 2014; Lehne, Shapiro, and Van den Eynde, 2018). Corruption increases the costs of providing public goods, either by awarding contracts to less efficient firms (Burguet and Che, 2004), or by funds disappearing altogether (Olken, 2006). However, due to its illegal and secretive nature, corruption is hard to detect and even harder to quantify. This is particularly true in cases of influence peddling and kickbacks, where buying influence is distinguishable only in a legal sense from legitimate activities such as lobbying or marketing (Goldberg, 2017).

This paper focuses on a setting where the distinction between legitimate marketing and influence peddling is especially hard to make: sponsorships awarded by pharmaceutical firms to doctors. In particular, I study whether private sponsorships such as conference expenses and speaking fees act as

[1] In 2017, more than 70% of health expenditure in OECD represented spending covered by government or compulsory insurance schemes. The figures are the result of own calculations using OECD's Health expenditure and financing data. Furthermore, global health expenditures are expected to more than double in the next 20 years (Dieleman et al., 2016).

kickbacks to doctors in public hospitals in exchange for procurement contracts. I combine a variety of administrative sources to create a unique dataset of 965,662 procurement contracts awarded by Romanian public hospitals in 2015-2016 and link them to the sponsorships awarded by pharmaceutical companies to doctors working in public hospitals. Sponsorships consist mostly of companies paying doctors' conference expenses and awarding them speaking fees for various events.

I document a timing effect: within three months of a sponsorship, there is a 4-5 percentage point increase in the probability of sponsoring firms winning procurement contracts. Also, a 1000 euro increase in sponsorships is linked to a 1 percentage point increase in the probability of obtaining a contract in the next three months. Furthermore, conditional on a procurement contract being signed, procurement contracts linked to sponsorships are 11% larger than those not linked to sponsorships.

Establishing such an association between private sponsorship and procurement is an important finding, as it prompts questions regarding its nature and the effectiveness of the procurement regulation. There are two main explanations for the timing effect: marketing and kickbacks. The marketing explanation consists of pharmaceutical firms investing in doctors' human capital by sponsoring their attendance to various professional events. During the events, doctors obtain information about new products and technologies, including those of the sponsoring firm, which increases the probability that the sponsoring firm receives a procurement contract from the hospital where the doctor works. A second explanation is that sponsorships act as kickbacks to doctors: in exchange for sponsorships, doctors use their influence to manipulate the hospital procurement process for the benefit of their sponsors. Thus, the second aim of the paper is to provide some steps towards disentangling those two explanations.

In order to quantitatively assess the importance of the two explanations, I exploit heterogeneity in the doctors who receive the sponsorships. Cole and Tran (2011) document a case study involving a bribe-paying pharmaceutical firm and identify two types of hospital staff that need to be bribed: the management staff and the prescribing doctors. I follow their example by differentiating between doctors in management position and regular doctors. The former are legally endowed with decision powers in hospital procurement and decide how the hospital allocates resoures, while the latter have at most an advisory role. If sponsorships act as kickbacks, the link between sponsorships and procurement should be stronger in the case of doctors in management positions as opposed to regular doctors.

I collect administrative data on public servants to identify sponsorships to doctors who hold management positions in public hospitals. Conditional on a contract being signed, a 1000 euro increase in the value of sponsorships is linked to a 10.41% increase in the value of the contract if the sponsorship was awarded to a doctor in management and a 9% increase if the sponsorship was awarded to a regular doctor. On the extensive margin, a 1000 euro increase in management sponsorships is associated with a 3.7 percentage point increase in the probability of getting a contract in the next quarter, while for regular sponsorships the increase is only 1.1 p.p. The difference stems mostly from direct contracts, which are awarded without any tender and are the least transparent of all procurement contracts. While the increased probability of obtaining a contract and the larger value of contracts associated with sponsorships is consistent with influence peddling or kickbacks, the similarity of this association between sponsorships in management positions and sponsorships to regular doctors is surprising.

An important argument against the marketing explanation is rooted in the institutional setting. Romanian procurement law restricts the legal avenues allowing pharmaceutical firms to influence hospitals' procurement process and has a wide definition for conflicts of interest. Also, pharmaceutical firms active in Romania are not allowed to condition prescription behaviour with sponsorships and they are legally required to send Romanian authorities all informational materials of their promotional events, including those occuring during scientific conferences. Taken together, these rules severely limit the scope of legitimate marketing by pharmaceutical firms. Furthermore, the majority of all procurement contracts (88.92%) and of those linked to sponsorships (93.44%) cover products that were bought before, in the past three years. This would suggest hospitals are relatively informed about the market, so there is reason to believe the scope of an informational marketing channel is limited.

Finally, I investigate whether contracts linked to sponsorships are less transparent, a situation also consistent with sponsorships acting as kickbacks. Following the literature on red flags in procurement (Fazekas and Toth, 2016), I use several measures of transparency: length of procurement, number of bidders, indicators for single bidder. Contracts linked to sponsorships are associated with shorter procurement times: being linked to a management sponsorship is associated with a decrease of 4 days in the time between announcement and signing. In contrast, the association between non-management sponsorships and the length of the procurement process is insignificant. However, there seems to be no link between sponsorships and the probability of having a single bidder or on the average number of bidders.

This paper makes several contributions. First, the Romanian institutional setting offers a unique chance to study corruption and kickbacks in a setting where high-quality formal legal institutions are in place. On the one hand, as a full-fledged EU member since 2007, Romania has a developed legal system and is bound by EU transparency regulation, regularly publishing data on procurement and spending. On the other hand, perceived corruption is pervasive especially in the healthcare sector, despite a legal system that formally limits conflicts of interest and bribes.[2]

While the existence of ties between politicians and and the private sector has been extensively studied both in economics (Asher and Novosad, 2017) and in political science (Desai and Olofsgård, 2011), the importance of links between other types of public bodies and firms has received much less empirical attention. Recent work by Tabakovic and Wollmann (2018) showed evidence that the revolving door between the private sector and patent regulators in the US is consistent with regulatory capture. The present paper is in a similar spirit: while outright quantitative identification of corrupt behaviour in complex institutions is virtually impossible to obtain, the question of the extent of private influence in public institutions is important enough to encourage investigation. In the medical literature, DeJong et al. (2016) showed that pharmaceutical industry-sponsored meals influence doctors' prescription behaviour and Larkin et al. (2017) showed modest effects on prescription behaviour after limiting detailing policies, while Abraham (2002) documented the ways in which pharmaceutical companies act as political players. However, to my knowledge, there are no studies linking the pharmaceutical sponsorships to procurement, which is more tightly regulated than individual prescription behaviour. The present paper also includes more diverse types of sponsorships, such as conference expenses and speaking fees and specifically differentiates between sponsorships to doctors in management and regular doctors.

A wide array of topics have recently been studied using procurement data, from waste in public services (Bandiera, Prat, and Valletti, 2009; Palguta and Pertold, 2017), to the effects of regulation on efficiency (Coviello and Mariniello, 2014) and transparency. An emerging strand of literature, pioneered by Mironov and Zhuravskaya (2016) aims to measure instititional quality and corruption at the local level by relating the distribution of procurement contracts to illicit financial flows. The current paper adds to this emerging field, linking legal financial flows (pharmaceutical sponsorships) to the distribution of procurement contracts.

Corruption is notoriously hard to identify due to its illegal nature, which is why early literature focused mostly on corruption perceptions (Reinikka and Svensson, 2006; Fisman and Gatti, 2002) or instrumental variables (Fisman and Miguel, 2007). However, the increase in data availability has made it possible to identify corruption in public spending using audits (Reinikka and Svensson, 2004; Olken, 2006; Ferraz and Finan, 2008) or administrative spending data (Lehne, Shapiro, and Van den

[2]Transparency International's Corruption Perception Index (Transparency International, 2017) routinely scores Romania at the bottom of the EU member states; in 2017, it scored 57th out of 180 countries surveyed. In November 2017, Laura-Codruta Kovesi, the head of the Romanian anti-corruption agency (DNA), declared: "... the bribe is paid in cash, before signing the contract and represents a certain percentage of the value of the contract. In healthcare, the bribe is about 20%, in IT 10% and in infrastructure beween 2% and 5%.

Eynde, 2018). This paper contributes to this literature by documenting a timing effect consistent with kickbacks.

The present paper is structured as follows. Section 4.2 presents the institutional setting of hospital procurement and pharmaceutical sponsorships. Section 4.3 presents the data collection process and describes the final samples, while Section 4.4 describes the empirical strategy. Section 4.5 provides an overview of the results, while Section 4.7 concludes.

4.2 Institutional setting

In this section, I present the institutional setting in which Romanian public hospitals procure their supplies. I cover three layers of institutions that are meant to improve transparency and avoid conflicts of interest: public procurement law, pharmaceutical sponsorship law and the practice in Romanian hospitals. I also present two very recent case studies, one exemplifying foul play in procurement, the other in pharmaceutical sponsorships.

4.2.1 Procurement rules

The Romanian law provides a plethora of public procurement procedures, based on the complexity of the contracts to be awarded. I focus on the four main types of procedures used in the dataset for the purchase of goods and services: direct procurement, invitations, negotiations and auctions. In what follows, I will use the words contracting authority and ”public body” interchangeably, to mean the hospital which is doing the procurement.

Before organising a procurement procedure, the contracting authority needs to estimate the value of the contract which will be awarded. The estimation is fully under the control of the institution, as the legislation provides solely loose guidelines. Based on the estimated value, the institution can choose the awarding procedure: open auctions are the gold standard, but smaller contracts can be awarded through simplified procedures such as invitations or negotiations. Each simplified procedure is only available below certain thresholds, so high-value contracts are usually awarded using open auctions.

Direct procurement is the simplest procedure. Since June 2013, it is applicable only to contracts below 30,000 euro (RON 132,519 in 2017). Direct procurement requires no auction and no announcement: the institution can choose its preferred supplier. Following Romanian law, I will use the word tender to express any procurement procedures that is not direct.

Invitations are simplified procurement procedures, applicable to contracts below 134,000 euro (RON 600,129 in 2017). In the first phase, the contracting authority announces the opening of the procedure, describing the contract to be awarded and the conditions to be fulfilled by firms. Firms send their files, applying for the right to make an offer and the authority selects its preferred candidates. In order to proceed to the second phase, the contracting authority needs to select a minimum of three candidates that it invites to make an offer. In the second phase, the firms send their offers and the contracting authority chooses the best fit according to the conditions set in the announcement.

Open and limited auctions are the most transparent procurement procedures: they are compulsory for contracts above 134,000 euro (RON 600,129 in 2017), but can also be used for smaller contracts. Open auctions have solely one phase, where the contracting authority publishes the announcement setting the contract conditions and deadlines. Any firm can make an offer within the established timeframe, after which the authority makes its choice. Limited auctions have those phases, one for selecting the candidates (similar to invitations) and one where firms bid for the contract.

Contracts above 134,000 euro need to be published in the Journal of the European Union and have strict rules regarding timelines and deadlines, while contracts below that threshold need to be published solely on the national portal and have laxer rules for participation.

Negotiations without participation announcements are procedures used when an auction has failed to supply a successful bid. The contracting authorities negotiate with the selected firms from the previous auction, in order to obtain a better contract than during the first procurement procedure.

It is explicitly forbidden to split large contracts into smaller ones in order to avoid the transparency thresholds. Conversely, splitting contracts into lots is particularly encouraged, as it allows SME's to compete for government contracts. If a large contract is awarded without being split into lots, the National Procurement Agency (ANAP) might classify the procurement as suspicious.

The law is purposefully vague regarding conflicts of interest, in order to capture a wide array of scenarios. It defines the conflict of interest as any situation where the employees of the public body who are involved in the procurement procedure have an economic, financial or personal interest that could compromise their impartiality or independence in the context of the tender. Furthermore, the law specifically forbids the winning firm from signing any commercial contracts related to the tender with the persons involved in the tender. This limitation is valid for 12 months after signing the contract and is aimed at avoiding the revolving door between public and private sector.

4.2.2 Procurement in hospitals

In order for a doctor's order to become a tender, it needs to pass through several layers of approvals. First, doctors need to explain the need for a certain substance: they need to file a needs report, which then gets approved by hospital management. Then, a team formed of various specialists create the technical requirements for the tender. Once the technical requirements have been set, they are approved by management and sent to the procurement office.

The procurement office is in charge of organising the tender: estimating the value of the tender, choosing the type of procedure, the deadlines and the attribution criteria. In theory, the attribution criteria can be one of the following: lowest price, best quality to price ratio, lowest cost and offer most advantageous from an economic point of view. However, in practice, procurement officers almost always choose the lowest price as the tender criterium.[3] The reason is that any of the other attribution criteria would be vulnerable to contestation, since it is hard to defend the public body's judgment of quality or economic sense. Thus, procurement officers strongly prefer any kind of quality conditions to be set in the technical requirements. This can also be seen in the data in Table 4.1: an overwhelming 99% of the contracts attributed through a tender had the attribution criteria set to "lowest price".

Table 4.1: Type of tender criteria

Name	N	% Total
Lowest price	28,124	99.68
Best quality	48	0.17
Lowest cost	28	0.10
Offer most advantageous from an economic perspective	14	0.05
Total	28,214	

After the tender is set up, it is posted on the national procurement website (`www.e-licitatie.ro`), together with the deadlines for sending an offer. Once the deadline passes, the office decides which offers satisfy the technical requirements and chooses the offer with the lowest price.

Although it is forbidden by the law, buyers could theoretically influence the chances of preferred suppliers obtaining contracts through modifications of the technical requirements, which are hard to identify ex-ante. According to the Romanian anti-corruption prosecutor's office, this is also one of the main channels through which corrupt officials can manipulate tenders (DNA, 2017a). An illustrative example is a hospital manager convicted of manipulating technical requirements to favor two companies in exchange for 10 and 20% of the contract values, respectively (DNA, 2017b). This added up to 140,000 euros for for the period 2010-2014, or ten times the gross yearly salary of a specialist physician.

[3] I would like to thank the procurement officer in a large Romanian public hospital for taking the time to explain all the practical procedures.

4.2.3 Pharmaceutical sponsorships

Since 2014, all sponsorships offered by pharmaceutical companies need to be reported to the Romanian National Agency of Medicines and Medical Equipment. According to pharmaceutical companies, sponsorships are intended to benefit patients by supporting doctors' continued education.

The law specifically forbids any kind of link between sponsorships awarded by pharmaceutical companies and the medication prescribed by sponsored doctors. Companies are forbidden from sponsoring scientific activities such as congresses and research projects where specific medication is advertised and are required to declare in advance to the Romanian authorities the topics, methods and materials to be used during conferences. In the case of sponsorships from pharma companies directly to hospitals, it is specifically forbidden that the sponsorship refer to any specific medication. Thus, the only legal reason for sponsoring a hospital is the generic "benefit of the patient".

Furthermore, it is specifically forbidden that companies condition doctors' prescription behaviour. Thus, they cannot offer any kind of gift or other benefit to doctors in exchange for prescribing a given drug. It is, however, allowed to offer doctors promotional objects of maximum 150 RON (around 30 euros), as long as they are relevant to practicing medicine.

Despite being specifically forbidden by law, there is reason to believe that some pharmaceutical companies have created incentives that encourage doctors to prescribe their drugs, which would go against the spirit of the law. An article published in 2018 by the investigative journalists of RISE Project (RISE Project, 2018) used internal documents of a large pharmaceutical company to show such an incentive scheme. For instance, doctors that had minimum five patients taking specific medication per year could be sponsored to go to national congresses, those that had 10-15 could go to European congresses, while those with an average of 25 per year could go overseas. Special rules applied to Key Opinion Leaders (KOL), who could obtain USD 200-300 in speaking fees per event.

4.3 Data

I used multiple administrative sources to create a dataset of all procurement contracts and the amount of sponsorships associated with each. Sponsorship data included the names of the doctors, but not their main affiliations. I used data on doctor affiliation and public servants' positions in order to match doctors to hospitals and differentiate between sponsorships to doctors in management ($lead$ sponsorships) and regular doctors ($other$ sponsorships). Then, I matched the daily firm-hospital sponsorship data with daily firm-hospital procurement data using the contracts dates. In this Section, I give details about the matching process and the assumptions underlying it.

4.3.1 Main variables: linking procurement to sponsorships

Throughout the analysis, I will use the following definition of a link: a procurement contract signed between firm f and hospital h in month m and year y is linked to a sponsorship if a sponsorship from firm f to hospital m occurred less than 3 months before the signing date. Legally, the largest minimum time elapsed between announcing a tender and the deadline for making an offer is 35 days. This gives minimum 2 months for the preparation of the tender and establishing a winner, so such a tight window should leave little time for other external factors to interfere with the procurement process, but enough time for sponsorships to affect procurement.

I use two main sources of heterogeneity in the sponsorship contracts. On the one hand, I check whether the timing of a sponsorship contract is related to the timing of a procurement contract. This requires a dummy variable Spons_yes_type$_{f,h,m,y}$, which is 1 if there was a sponsorship contract of $type$ (leadership sponsorship, other sponsorship or any sponsorship) between firm f and hospital h less than three months before month m, year y. On the other hand, I use the variation in the amount of sponsorships, which is captured by variable Spons_EUR1000_type$_{f,h,m,y}$: the amount (in thousands of euros) of $type$ sponsorships flowing from firm f to hospital h three months before month m and year y.

The original data was in RON, so for ease of exposition I transformed it into euro using the average exchange rate of €1 = RON 4.4679.

4.3.2 Data sources

Sponsorships, doctor affiliation and management positions

The sponsorship data was obtained from the website of the official Romanian National Agency of Medicines and Medical Equipment[4] (ANM). The original data included 128,816 sponsorship contracts for the period 2015-2016. The data included inter alia the names of the recipients, their specialization (in freeform), the value and date of the contracts, the name of the sponosor and a free-form description of the contract. It also included all sponsorship contracts, including those awarded to institutions such as patient associations, nurses, GP's, doctors, non-medical personnel etc.

The first step in using this data was to link the recipients of the sponsorships to the institutions where they worked. The raw sponsorship data provided only the names of the beneficiaries of each sponsorship, without their institutional affiliations. I decided to focus only on recipients that were medical doctors, as public data on nurses' activities or other personnel (researchers, professors, support staff) is scarce. I also removed all GP's, as they are considered freelancers and are not linked

[4]The website can be found at `https://www.anm.ro/`.

to public hospitals. I used scraped data from the official registry of the Romanian National Doctor's College[5] (CM) in order to match doctors to public hospitals. This was done using vectorial decomposition with a threshold of 1, combined with manual matching using internet searches. The latter was only used if I could find a clear match using a doctor's name, specialisation and something in the address that could pinpoint the hospital. The conservative choice of the matching algorithm threshold was done in order to limit misplacing doctors as much as possible. The CM data included women's previous names, which were only used in case of failing to match the current name. In all cases, I kept solely persons identified as doctors in the sponsorship sample, either by their "Dr." title or by their job description.

A small percentage of the sponsorships were given directly to public hospitals, rather than to specific persons in those hospitals. These sponsorship contracts were deemed institutional sponsorships, in order to differentiate them from sponsorships given to persons within the hospitals. They include medicine and equipment donations, renovations etc and tend to be larger than the standard conference or professional association fees. I identified institutional sponsorships using the name (those that had hospital, association etc in their name) and matched them to the unique hospital ID's. However, for the purpose of the main analysis, I will ignore institutional sponsorships: they will only be included as a robustness check.

In order to link doctors to management positions, I used data from the Romanian National Integrity Agency.[6] The portal includes all the asset declarations for Romanian public servants who hold management positions. Asset declarations are legally binding and need to be filed at least once a year and upon beginning a new management position. The metadata (name, county, date, position, but not institution) is available on the Agency's data portal. Regrettably, hospitals are clustered together under the categories "Health Ministry - other institutions" and "Other institutions subordinated to the county councils". After downloading all the metadata from these categories and homogenizing the county names, I matched the sponsorship data for the doctors with the asset declarations, using the name and the county. I identified leadership positions as any of the following in the year of obtaining the sponsorship: director, president, chief, manager, coordinator, board member. I used very stringent matching criteria for linking the asset declaration data with the sponsorship data: the name and county of the person had to be exactly the same and only the word order could vary (vectorial decomposition with a threshold of 1). This opens the possibility that some managers might have been excluded, due to spelling errors in the names.

[5] The website was the National Doctor's Registry and can be found at `https://regmed.cmr.ro/`.

[6] The address can be found at `http://declaratii.integritate.eu/`

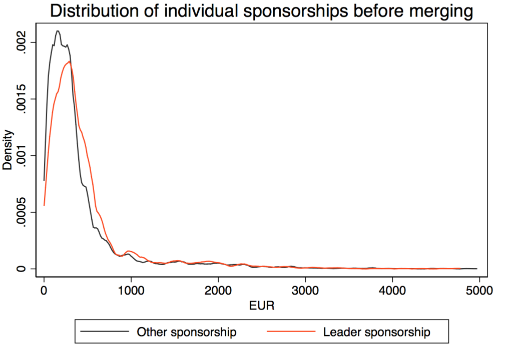
As it can be seen in Table 4.2, the matched dataset included 22,277 sponsorship contracts over the period 2015-2016. Of these, 3,397 were offered to doctors in management positions and the rest were offered to doctors who did not have management positions (18,880). Throughout the rest of the analysis, a ”lead” sponsorship will denote a sponsorship given to someone in management, whereas a sponsorship of the type ”other” will denote a sponsorship given to any other doctor in the hospital. The dataset included 94 pharmaceutical companies and 313 hospitals. The contracts went to 4040 doctors, of whom 373 had management positions. In total, the leaders cover 155 hospitals. Figure

Table 4.2: Descriptive statistics, raw sponsorships (in euro)

Sponsorship type	N	Mean	Std.dev	Median	Max
All	22,277	426.7886	557.3152	268.5826	11184.45
Lead	3,397	480.4739	576.191	322.2991	11184.45
Other	18,880	417.1292	553.3137	263.4594	8239.979

4.1 shows an overview of the distribution of sponsorships. The bulk of the sponsorships (99.99%) are under 5,000 euro, with sponsorships to managers tending to be slightly larger.

Figure 4.1: The distribution of raw sponsorships



4.3.3 Procurement data

The procurement data (contracts, announcements, subsequent contracts etc) was downloaded from the official open data portal of the Romanian government (`data.gov.ro`). The data is curated by the the National Public Procurement Agency (ANAP) and updated every three months. The raw data consisted of the universe of procurement contracts in Romania, of which I cleaned solely those signed by the public hospitals in Romania.

The sample was limited to the period January 2015 - December 2016, in order to match the sponsorship data. Public hospitals were identified using public data on the Ministry of Health's portal on hospital expenses (www.monitorizarecheltuieli.ms.ro) and the list of hospitals bound by an ethics council, which I found on the government's data portal. I excluded clinics, private healthcare providers and GP practices, as the focus of this paper is solely on public hospitals. The final list of hospitals consisted of 400 entities. This left a total of 376 hospitals in the matched sponsorship-procurement sample. Some hospitals were shut down or merged in the period 2008-2014, which explains the smaller number of hospitals present in the procurement data in 2015-2016.

The data was limited to include solely contracts between hospitals and firms who have sold minimum one product whose CPV code starts with 33 (medical equipment, pharmaceuticals and personal care products). This limitation ensured that I would focus only on firms connected to the medical market, excluding firms specialised in other sectors which are not related to sponsorships (industry, legal services, IT etc).

I only took into account the contracts that were for goods and services. Works contracts have different transparency thresholds and rules, tend to be much larger than contracts for goods and services and are part of a completely different market, so I decided to remove them from the sample. This left 965,662 contracts that included goods and services.

An important note is that the final sample includes only the initial tender contracts upon the date of the signing of the contract. I excluded subsequent contracts, as the data quality was uncertain and it was not clear which subsequent contracts were linked to which procedure. Furthermore, subsequent contracts could be signed years after a procedure, which would be hard to detect using only the sample 2015-2016. Focusing solely on the initial contract value mitigates this issue. Further details about the consolidation of lots into contracts and assumptions regarding missing values can be found in the Appendix.

4.3.4 Descriptive statistics: main sample

Number of contracts linked to sponsorships

The final sample consists of 965,662 contracts. Table 4.3 contains a breakdown of each type of contract, by links to sponsorships and by procedure type. It can be observed that the vast majority (97%) of procurement contracts are actually direct contracts, which are not awarded using any kind of tender. Furthermore, a fifth (22.6%) of the hospitals in the sample have at least one contract linked to a sponsorship. Furthermore, as it can be seen in Table 4.3, a very small percentage (0.71%) of the firms

Table 4.3: Number of entities with at least one linked contract

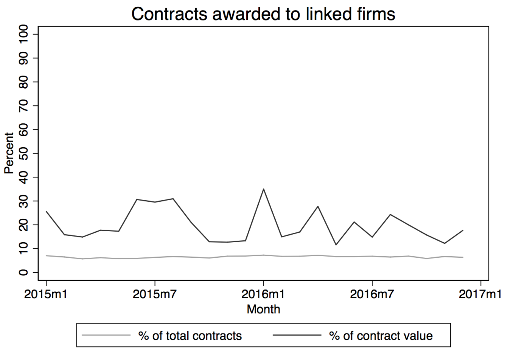
	Linked to a sponsorship in the past 3 months?					
Contracts	**Yes**		**No**		**Total**	
	N	% Total	N	% Total	N	% Total
Direct procurement	2737	0.3	934708	99.7	937445	100.0
Participation invitation	53	0.9	5953	99.1	6006	100.0
Open auction	340	2.2	15388	97.8	15728	100.0
Negotiation (no ann.)	149	2.3	6331	97.7	6480	100.0
Total	3279	0.3	962380	99.7	965659	100.0
Firms						
	15	0.71		99.29	2105	100
Hospitals						
	85	22.6	291	77.4	376	100

that were awarded procurement contracts in 2015-2016 also have contracts linked to sponsorships.[7] Figure 4.2 shows that these 15 firms have won a significant percentage of total procurement: they account for 20.27% of the total value of procurement contracts that occurred between 2015 and 2016. Table 4.14 shows that 7 out of the 15 firms that have linked contracts are in the top 30 of the firms with the largest value of total procurement contracts in the sample. The total value of procurement contracts between 2015 and 2016 was 2,970 million euros, of which 3.93% (117 million euros) was linked to sponsorships.

This suggests that although there are very few firms that have linked contracts, they are important players in the procurement market and a significant portion of the sponsorships market.

[7]There are also 5 other firms who have both sponsorship and procurement contracts, but those contracts cannot be linked using the window of three months before the signing date. Thus, these firms will be treated as not linked for the remainder of the paper.

Figure 4.2: Contracts going to linked firms



Note: Linked firms are firms with minimum one linked contract in 2015-2016.

Distribution of contract values

Tables 4.4 and 4.5 show an overview of the main variables of interest:[8] the value of sponsorships in the past three months and the contract value. The majority of contracts (92.94%) are under 1,000 euro, while contracts above 100,000 euro represent less than 1% of the sample. Unsurprisingly, the sponsorship variables follow the distribution of the raw sponsorships: sponsorships to management are fewer, but slightly larger than sponsorships awarded to regular doctors: the mean stands at 596 euros for the former and 544 euros for the latter.

Table 4.4: Descriptive statistics, main explanatory variables (in thousands of euro)

Variable name	N > 0	Mean	Std.dev	Median	Max
Spons_EUR1000	3,279	0.5992	1.1614	0.2284	16.6128
Spons_EUR1000_other	3,062	0.5437	1.0789	0.2140	16.6128
Spons_EUR1000_lead	503	0.5959	0.9456	0.2283	7.3646

Table 4.5: Distribution of contract values

Contract Value (euro)	N	% of total
< 1000	897484	92.94
1,000 – 10,000	53916	5.58
100,000 – 1,000,000	3,214	0.33
> 1,000,000	529	0.55

[8]More detailed descriptive statistics can be found in the Appendix, in Table 4.13 and Figures 4.4.

Repeated purchases

As it can be seen in Table 4.6, more than 80% of the goods and services purchased in the procurement dataset belong to a category (CPV code) that was purchased in the previous years, with the percentage approaching 90% when the past four years of procurement are taken into account. Thus, while it is possible that pharmaceutical companies use sponsorships to organize informational events for doctors and persuade them to purchase their products, the extent of this channel is likely to be limited.

Table 4.6: Previous purchases

Description	# Total	% Total	# Linked	# Not linked
Product purchased past year	804,502	83.31	2,839	801,663
Product purchased past 2 years	842,169	87.21	2,995	839,174
Product purchased past 3 years	855,604	88.60	3,064	855,604
Product purchased past 4 years	866,575	89.74	3,108	866,575

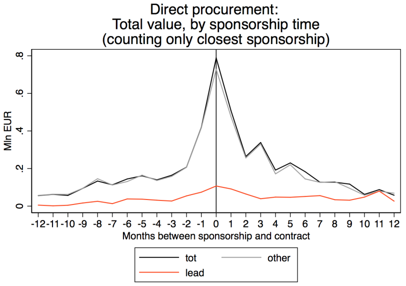
4.4 Empirical strategy

4.4.1 Pseudo-event study

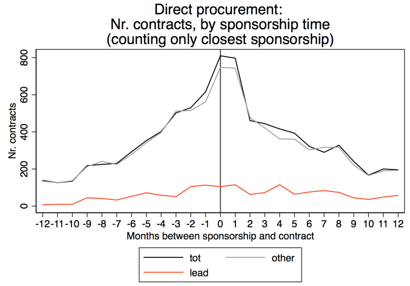
A natural first step towards establishing whether there is an relationship between sponsorships and procurement contract is an event study around the time of the sponsorship. Although the sponsorship decision is likely endogenous, such an exercise can be informative and provide circumstantial evidence of a link between sponsorships and procurement, regardless of the direction of the relationship.

Figure 4.3 provides such an event study exercise using the number and total value of procurement contracts. In order to establish the timing between a contract and a sponsorship, I chose only the closest sponsorship to the date of the signing. Thus, if a firm gave a hospital sponsorships in February and May and signed a procurement contract in April, that contract would be both one month before the May sponsorship and two months after the February sponsorship. In Figure 4.3, I assign such a contract value -1, as the absolute value is the smallest in the case of the May sponsorship. I do a similar exercise for all the procurement contracts linked to a sponsorship and sum up over the relative timing categories. As it can be observed in Figure 4.3, there seems to be a relative timing effect in the number of contracts that get signed: for all the contracts that are possibly linked to a sponsorship one year before and after the signing date, the bulk of the contracts seem to occur very close to the date of the sponsorship. A similar pattern can be found for the value of contracts, but the variability is much higher in the case of large tender contracts.

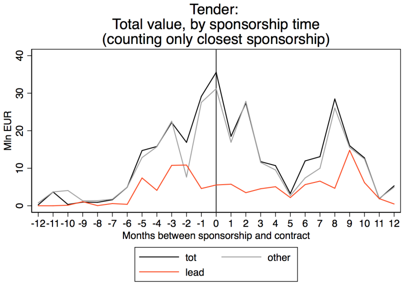
Figure 4.3: Timing of sponsorships and procurement contracts



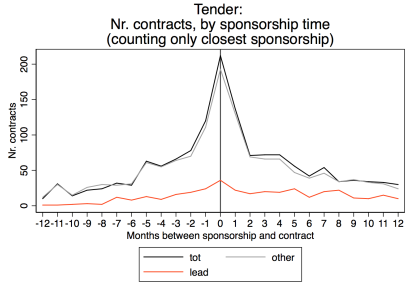
(a) Value of direct contracts



(b) Number of direct contracts



(c) Value of tender contracts



(d) Number of tender contracts

Note: the timing was calculated using the closest sponsorship (in terms of absolute value of the difference between month of signing the sponsorship and the month of signing the procurement contract).

4.4.2 Baseline estimations: intensive margin

The first task is to study whether contracts that are linked to procurement sponsorships have a different value than those that are not. Although this is a question about the intensive margin of the sponsorship, there are several reasons to study it before the extensive margin. First, it shows the link between sponsorships and procurement, conditional on contracts being won, so it can be answered using the variation in contracts that are actually signed. Second, estimating the intensive margin of the relationship helps set up the framework that will also be used to estimate the extensive margin.

The unit of analysis is the procurement contract between firm f and hospital h and it includes the value of all lots associated with this contract. I estimate the following equation:

$$\begin{aligned}\log V_{f,h,p,m,y} =& \beta_1 SponsPast3Months_{f,h,p,m,y} \\ &+ \sum_i \gamma_i Procedure^i_{f,h,p,m,y} + \eta_f + \eta_h + \eta_p + \eta_m + \eta_y + \varepsilon_{f,h,p,m,y}\end{aligned} \tag{4.1}$$

The left-hand side ($\log V_{f,h,p,m,y}$) is the log value of the procurement contract and the main variable of interest is $SponsPast3Months_{f,h,p,m,y}$: the total value of sponsorships in the three months preceding the procurement contract. The rich data allows for controlling for a large range of fixed effects (firm f, hospital h, product p, month m and year y) and for procedure type, so this should limit other channels that link sponsorships to procurement contracts.

The parameter of interest is β_1, which shows whether contracts related to sponsorships are larger than contracts not related to sponsorships. The parameter β_1 can be interpreted as the strength of the link between sponsorship and contract value. Conditional on a number of factors including time-, firm- and hospital- specific effects, there is no a priori reason why contracts linked to sponsorships should be any different than those not linked. Thus, a statistically significant β_1 would be consistent with influence peddling by the recipients of the sponsorships who lobby for their sponsoring firms within their hospitals. This leads to Hypothesis 1:

Hypothesis 1. $\beta_1 > 0$*: Procurement contracts that are linked to sponsorships have larger values than contracts that are not linked to sponsorships.*

After establishing whether there is indeed an association between the procurement value and having been sponsored, I study the heterogeneity of this link across different types of sponsorships.

In other words, I estimate the following equation:

$$\log V_{f,h,p,m,y} = \beta_2 LeadSponsPast3Months_{f,h,p,m,y} + \beta_3 OtherSponsPast3Months_{f,h,p,m,y} + \sum_i \gamma_i Procedure^i_{f,h,p,m,y} + \zeta_f + \zeta_h + \zeta_p + \zeta_m + \zeta_y + \varepsilon_{f,h,p,m,y} \tag{4.2}$$

The left-hand side ($\log V_{f,h,p,m,y}$) is the log value of the procurement contract and is the same as in equation (4.1). The sponsorship variable $SponsPast3Months_{f,h,p,m,y}$ from equation (4.1) has been split into two types of sponsorships: sponsorships to management (LeadSponsPast3Months) and sponsorships to other doctors in the hospital (OtherSponsPast3Months), in order to capture the different decision powers of the sponsorship recipients.

Again, there is no a priori reason to believe contracts linked to sponsorships would be any different than those not linked to sponsorships, so a positive association would provide evidence for foul play. Furthermore, if there were preferential treatment for the sponsors, the link between procurement value and leadership sponsorships would be larger than that between procurement values and sponsorships to other doctors, as this reflects the larger decision power of the hospital management compared to regular doctors. This leads to Hypothesis 2:

Hypothesis 2.

1. $\beta_2 > 0$, $\beta_3 > 0$: *Sponsorships are positively related to the procurement contract values.*

2. $\beta_2 > \beta_3$: *The link between leadership sponsorships and procurement contract value is larger than that between other sponsorships and procurement contract value.*

4.5 Results

4.5.1 Intensive margin

Table 4.7 gives an overview of the main regression results. Sponsorship variables are measured using both dummies indicating whether a sponsorship took place in the three months prior to the signing of the contract (variables with suffix $\_yes$) and as the total value of sponsorship contracts, in thousands of euros (variables with suffix EUR1000). It can be observed from Table 4.7 that sponsorships and

Table 4.7: Intensive margin: main results

	(1) lnV	(2) lnV	(3) lnV	(4) lnV
Spons_yes	0.1100* (0.0582)			
Spons_yes_lead		0.0594 (0.1124)		
Spons_yes_other		0.1001 (0.0610)		
Spons_EUR1000			0.0921*** (0.0277)	
Spons_EUR1000_lead				0.1041** (0.0508)
Spons_EUR1000_other				0.0900*** (0.0313)
Observations	964952	964952	964952	964952
R^2	0.551	0.551	0.551	0.551
clustvar	spit_id	spit_id	spit_id	spit_id
Month f.e.	Yes	Yes	Yes	Yes
Year f.e.	Yes	Yes	Yes	Yes
Firm f.e.	Yes	Yes	Yes	Yes
Hospital f.e.	Yes	Yes	Yes	Yes
Product f.e	Yes	Yes	Yes	Yes

procurement contracts do seem to be related: procurement contracts linked to sponsorships are 11% larger than those not linked to sponsorships and a 1000 euro increase in sponsorships three months prior to a contract is associated with a 9.21% increase in the value of that contract. Thus, Hypothesis 1 can be confirmed on both counts: contracts linked to sponsorships are associated with larger values and the marginal effect of a sponsorship is positive and significant.

Due to the relatively small number of linked contracts in the sample, there is not enough variation in the existence of a sponsorship to precisely estimate the association between contract values and the two different types of sponsorships (lead and other). However, when the variation in the amount of sponsorship is taken into account, as it is in the fourth column of Table 4.7, precision improves significantly: both the sponsorships awarded to management and those awarded to other doctors are correlated with larger contract values. Although on the intensive margin, being linked to a management sponsorship seems to be correlated with a larger contract than being linked to a regular sponsorship (10% larger as opposed to 9% larger), the difference is of limited economic significance. Thus, while the first part of Hypothesis 2 can be confirmed by arguing that there is a positive association between sponsorships and contract values, the second part is harder answer unambiguously.

Tables 4.16 and 4.18 in the Appendix provide further detail into the association between sponsorships and procurement values by estimating non-linear specifications that use polynomials of second and third degree. Since the coefficients of the higher-order polynomials are insignificant, it can be concluded that the association between sponsorships and contract values is best estimated by the log-lin specification in Table 4.7.

4.5.2 Extensive margin

Balanced panel: descriptive statistics

Since the dataset used for estimating the link between contracts and sponsorships along the intensive margin consists only of contracts that have been won, no counterfactuals are readily observable. Consequently, in order to estimate the relationship between the likelihood of getting a procurement contract and pharmaceutical sponsorships, one needs to exploit variability in getting a contract. To this end, I create a monthly balanced panel of firm-hospital pairs, where I record whether a sponsorship or a procurement contract occurred in month m between firm f and hospital h. Since firms needs to register on the procurement portal prior to obtaining a contract, the set of firms that have sold anything in 2015-2016 would be a good approximation for the universe of firms who could have received such a contract. However, due to memory and computational limitations, I could not create a balanced panel that would include product codes: such a panel would have quickly run into the curse of dimensionality, since there are 4,070 different CPV codes in the dataset.

The balanced panel includes 1,203,096 observations of hospital-firm pairs, observed through the 24 months between January 2015 and December 2016. Table 4.8 presents the descriptive statistics of the balanced panel created for this purpose: 238,198 observations with a procurement contract and 3,686 observations with minimum one sponsorship contract.

Table 4.8: Balanced panel descriptive statistics

Statistic	Value
No. firm-hospital-time pairs	1,203,096
No. firm-hospital pairs	50,129
No. obs with min. 1 procurement contract	238,198
No. obs with min. 1 direct contract	23,920
No. obs with min. 1 tender contract	23,920
No. obs with min. 1 sponsorship	3,686
No. obs with min. other sponsorship	3,448
No. obs with min. leadership sponsorship	653
No. months	24

Balanced panel: empirical strategy

In this section, I establish the extensive margin of the relationship between sponsorships and procurement, i.e., whether sponsorships are related to a higher likelihood of obtaining procurement contracts in the quarter after they are signed. In order to obtain an answer. I estimate equations similar to the intensive margin:

$$\mathrm{I}(Contract)_{f,h,m,y} = \alpha_1 SponsPast3Months_{f,h,m,y} + \eta_f + \eta_h + \eta_m + \eta_y + \varepsilon_{f,h,m,y} \tag{4.3}$$

$\mathrm{I}(Contract)_{f,h,m,y}$ is an indicator variable which takes value 1 if a procurement contract took place between firm f and hospital h on month m in year y and is 0 otherwise, while $SponsPast3Months_{f,h,m,y}$ is the standard sponsorship variable. In order to estimate equation 4.3, I used a standard linear probability model with a rich level of fixed effects (firm f, hospital h, month m, year y), which should control for a large proportion of unobserved heterogeneity. Similarly to Hypothesis 1, a positive coefficient for sponsorships is consistent with influence peddling:

Hypothesis 3. $\alpha_1 > 0$*: Sponsorships less than 3 months before the contract date are correlated with an increase in the probability of receiving a procurement contract.*

Similar to Hypothesis 2, I also test for heterogeneous effects of different types of sponsorships. To do so, I estimate the following equation:

$$\begin{aligned}\mathrm{I}(Contract)_{f,h,m,y} = {} & \alpha_2 LeadSponsPast3Months_{f,h,m,y} + \alpha_3 OtherSponsPast3Months_{f,h,m,y} \\ & + \zeta_f + \zeta_h + \zeta_m + \zeta_y + \varepsilon_{f,h,m,y}\end{aligned} \tag{4.4}$$

In order to be consistent with influence peddling, the link between leadership sponsorships and the likelihood of getting a contract should be stronger than the link between other sponsorships and the likelihood of getting a procurement contract and both should be positive:

Hypothesis 4. *1.* $\alpha_2 > 0$, $\alpha_3 > 0$

2. $\alpha_2 > \alpha_3$*: The link between leadership sponsorships and the probability of getting a procurement contract is stronger than that between other sponsorships and the likelihood of obtaining a procurement contract.*

Balanced panel: extensive margin results

Table 4.9 provides the estimation results of the linear probability models described in equations (4.3) and (4.4). The variable $\mathrm{I}(Contract)_{f,h,m,y}$ on the left-hand side of the equations takes value 1 if a contract was signed in between firm f and hospital h in month m and year y. In order to take into account the possible heterogeneity in the association between sponsorships and controls for tenders and direct contracts, I use three different specifications for contracts: Any is 1 if any type of contract occurred, $Direct$ is 1 only if a direct contract occurred and $Tender$ is 1 if any type of contract but direct ones occurred.

It can be observed from Table 4.9 that sponsorships are associated with higher probabilities of a procurement contract occurring within the next three months, confirming Hypothesis 3. However, it is harder to establish whether Hypothesis 4 should be rejected: while the marginal effect of increasing management sponsorships with 1000 euros is associated with a significantly larger increase in the probability of obtaining a procurement contract than the marginal effect of increasing regular sponsorships (i.e, 3.7 percentage points increase as opposed to 1 p.p), contracts linked to management sponsorships are statistically indistinguishable from contracts that are not linked to any sponsorships. This lack of precision could once again be due to the limited amount of contracts which can be linked to management sponsorships, which increases the standard errors.

Table 4.9: Extensive margin results

	(1) Any	(2) Any	(3) Direct	(4) Direct	(5) Tender	(6) Tender
Spons_yes	0.0555***		0.0338***		0.0471***	
	(0.0107)		(0.0107)		(0.0077)	
Spons_yes_lead		0.0374		0.0253		0.0296*
		(0.0266)		(0.0249)		(0.0150)
Spons_yes_other		0.0520***		0.0327***		0.0427***
		(0.0106)		(0.0105)		(0.0083)
Observations	1203096	1203096	1203096	1203096	1203096	1203096
R^2	0.130	0.130	0.131	0.131	0.064	0.064
clustvar	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id

	(1) Any	(2) Any	(3) Direct	(4) Direct	(5) Tender	(6) Tender
Spons_EUR1000	0.014***		0.008*		0.008***	
	(0.005)		(0.004)		(0.002)	
Spons_EUR1000_lead		0.037*		0.033*		0.004
		(0.019)		(0.018)		(0.008)
Spons_EUR1000_other		0.010**		0.004		0.009***
		(0.005)		(0.004)		(0.002)
Observations	1203096	1203096	1203096	1203096	1203096	1203096
R^2	0.130	0.130	0.131	0.131	0.064	0.064
clustvar	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, firm

4.6 Association between red flags and sponsorships

The procurement literature (Fazekas and Toth, 2016) has identified a number of red flags that seem to correlate with corruption in public procurement. Examples such red flags are: small number of bids, unexpected changes in tender requirements and/or documentation, tight deadlines for bidders or highly complex tender documentation. Since no tender texts are available in my sample, I use three red flags which can be easily quantified: single-bid auctions, average number of bidders per firm-hospital-contract and procurement length. Since deadlines were not observable, I used the time elapsed between the contract announcement and its signing as a proxy for procurement length.

The estimation results can be seen in Tables 4.10 and 4.11. Since announcement dates are only observable for invitations and open auctions, the sample size is significantly reduced compared to the previous estimations. While there is no reason to believe that sponsorships are related with single-offer procurements, contracts related to management sponsorships are associated with auctions that are 4.5 days shorter than contracts not related to any sponsorship. Further robustness checks are provided in the Appendix (Tables 4.21, 4.22, 4.23, 4.24), both for auctions and for a sample containing both auctions and invitations. The conclusions are the same: sponsorships are not consistently related to less competitive procurements, but leadership sponsorships seem to be associated with shorter procedures.

Table 4.10: Are sponsorships associated with single-offer auctions?

	(1) SingleOffer	(2) SingleOffer	(3) SingleOffer	(4) SingleOffer
ContractValue1000	0.0000**	0.0000**	0.0000**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spons_yes	0.0214			
	(0.0131)			
Spons_yes_lead		0.0202		
		(0.0191)		
Spons_yes_other		0.0143		
		(0.0150)		
Spons_EUR1000			-0.0017	
			(0.0033)	
Spons_EUR1000_lead				0.0057
				(0.0071)
Spons_EUR1000_other				-0.0028
				(0.0037)
Observations	15439	15439	15439	15439
R^2	0.385	0.385	0.385	0.385
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, firm

Table 4.11: Are sponsorships associated with shorter auctions?

	(1) ProcLength	(2) ProcLength	(3) ProcLength	(4) ProcLength
ContractValue1000	-0.0004	-0.0004	-0.0004	-0.0004
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Spons_yes	-1.9939			
	(1.2724)			
Spons_yes_lead		-4.5764**		
		(1.8561)		
Spons_yes_other		-1.1402		
		(1.2892)		
Spons_EUR1000			-0.6303	
			(0.4624)	
Spons_EUR1000_lead				-2.8466**
				(1.1873)
Spons_EUR1000_other				-0.3138
				(0.4434)
Observations	15405	15405	15405	15405
R^2	0.619	0.619	0.619	0.619
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, firm

4.7 Conclusion

This paper has studied the association between pharmaceutical sponsorships and the procurement patterns of public hospitals. It documented a timing effect: sponsorships are associated with a 5% higher probability of the sponsoring firm obtaining a procurement contract within the next quarter. Furthermore, conditional on winning the procurement, sponsorships are associated with a 10% larger contract value. While selection issues are likely to be problematic on an econometric level, the relatively small increase in the effect of sponsorships on the probability of obtaining a contract limits their economic significance.

Next, I aimed to explain the reasons behind this timing effect. On the one hand, sponsorships could be related to procurement contracts because of legitimate marketing campaigns. The main mechanism put forward is that pharmaceutical firms sponsor doctors' attendance to informative scientific events where they are convinced of the value of the firms' products. However, this explanation is relatively unlikely due to the legal limits placed on pharmaceutical sponsorships and procurement regulation aimed at limiting conflicts of interest.

On the other hand, sponsorships could act as kickbacks to doctors either by incentivising the prescription of sponsors' products or by encouraging the manipulation of tenders in the sponsors' favor. To test whether the latter is more likely, I looked at whether the magnitude of the association between sponsorships and procurement is larger in the case of doctors holding management positions as opposed to regular doctors. However, since there was a relatively small sample of procurements linked to sponsorships, precise estimates of heterogeneous effects were difficult to obtain. On the intensive margin, the difference between sponsorships to doctors in management and those awarded to regular doctors is both statistically and economically insignificant. However, on the extensive margin, sponsorships to management seem to be associated more to an increase in the probability of awarding a contract, with the effect stemming mostly from direct contracts, which are the least transparent.

Finally, I investigated whether contracts linked to sponsorships are less transparent, which would be consistent with the kickbacks mechanism. Using red flags commonly used in procurement corruption studies, I investigated whether sponsorships are associated with more red flags. The difficulty in obtaining precise estimates of the heterogeneity in the association between sponsorships and procurement was also apparent when studying red flags. However, the red flags are known to the literature for their imprecision, which is why several of them need to be checked in order to reach a clear conclusion. In the current case, sponsorships do not seem to be clearly related to the number of offers a procurement contract is made, but sponsorships to management do seem to be related to tighter procedures.

Overall, it would seem that although there is clear reason to believe that there is a timing effect between sponsorships and procurement contracts, it is hard to establish that this is entirely due to corrupt behavior such as kickbacks. However, given the institutional setting and the association between management sponsorships and the likelihood of obtaining a procurement contract, this would still be a likely explanation. Further research is required to establish clear mechanisms.

Furthermore, given the importance of healthcare in public spending across the world, a natural question arises as to how do the links between sponsorships and procurement translate to patient outcomes. Future research should shed light on this association and further help in disentangling the information and corruption mechanisms.

4.A Data

4.A.1 Procurement dataset: further details

For procedures that are not direct procurement, there is usually a tender announcement that is recorded in the database. However, many of the announcement numbers for open tenders and invitations were missing. I updated the data on these missing announcement numbers using the list of tender calls available at `data.gov.ro`. However, for negotiations that do not need a participation announcement, no announcement ID was available. I created unique ID's for these tenders using the combination of public body, CPV code (EU-wide code for classifying all procurement products), contract date, contract number and contract type (procurement/framework). I also excluded contracts that should have had a tender ID but did not (invitations, open procurements etc), which left 965,662 contracts.

The procurement dataset included all the lots won by a specific company, which was shown as multiple contracts. For some of the contracts, the value of each lot was not given: only the final tender value was observed. In order to ensure the homogeneity of the data, I created a tender identifier using the estimated value, the hospital ID, CPV code and the tender date. Then, I created a total value for each company that won minimum one lot in that tender. If all lots had the same value, the value given was actually the total value of the contracts, so I approximated the total value of lots awarded to each company by dividing the total contract value by the number of winners.

If a company had won multiple lots, I collapsed them into a single value. I consolidated all procurement at the tender level (using the tender id): thus, if a tender organised by hospital h resulted in 3 lots worth 1000 euro being allocated to firm f_1 and 5 lots worth 2000 euro being allocated to firm f_2, the data would show only 2 observations: one worth 1000 euro organised between hospital h and firm f_1 and one contract worth 2000 euro between hospital h and firm f_2. This ensures that I take the total value earned by a firm from a specific tender. The final result is the "Contract Value" variable that will be used throughout this paper.

A small proportion of the contracts in the procurement dataset are framework contracts, as it can be seen in Table 4.12. Those contracts are not always cleanly recorded: there are multiple instances when a large tender is won by multiple firms, but the value of each hospital-firm pair is recorded as the total value of the contract[9]. For those contracts, I assumed that the lots were evenly spread among the firms and divided the repeating value of the contract by the number of firms which won it.

[9]This is quite clear, since the repeating value is usually close to the announcement value and there are many firms that won the contract. Adding up those values would result in a value that is multiple times larger than the estimated value

Table 4.12: Types of contracts

Type of contract	N	% Total
Direct procurement	937,445	97.07
Framework agreement	18,094	1.87
Regular contract	10,124	1.05

4.B Descriptive statistics

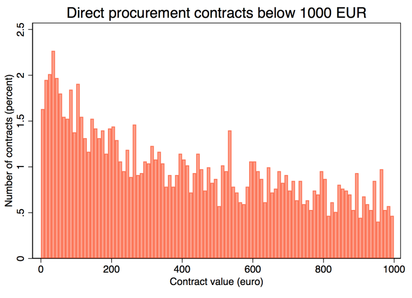
4.B.1 General descriptive statistics

Table 4.13: Descriptive statistics, main explanatory variables (in thousands of euro)

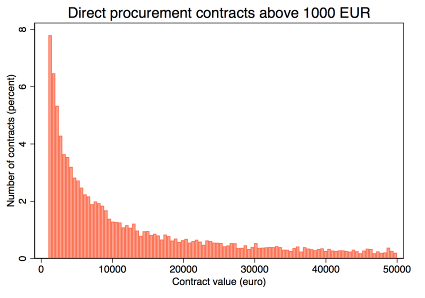
Variable name	$N > 0$	Mean	Std.dev	Median
Contract value	965,652	3.1971	95.2479	0.0616
Contract value|Spons_EUR1000 > 0	3,279	35.7739	312.132	0.2058
Contract value|Spons_EUR1000_other > 0	3,062	33.01608	292.3752	0.2053
Contract value|Spons_EUR1000_lead > 0	503	64.8998	529.3049	0.2619
Contract value|Spons_EUR1000 = 0	962,373	3.0861	93.6354	0.0616
Contract value|Spons_EUR1000_other = 0	962,590	3.1023	93.9486	0.0615
Contract value|Spons_EUR1000_lead = 0	965,149	3.1650	94.4943	0.6155

4.B.2 Distribution of procurement values

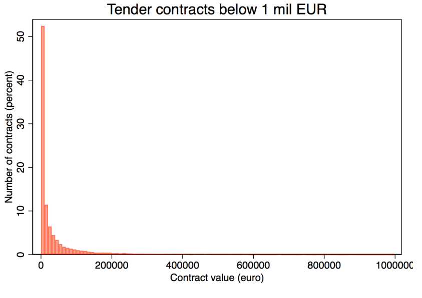
Figure 4.4: Distribution of procurement contract values



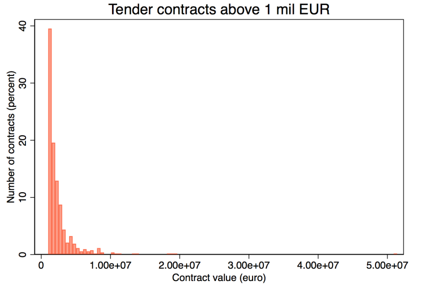
(a) Distribution of value of direct contracts below 1000 euro



(b) Distribution of value of direct contracts above 1000 euro

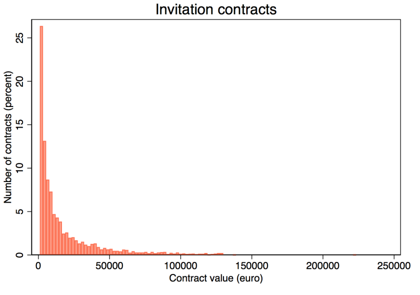


(c) Distribution of value of tenders below 1 mil euro

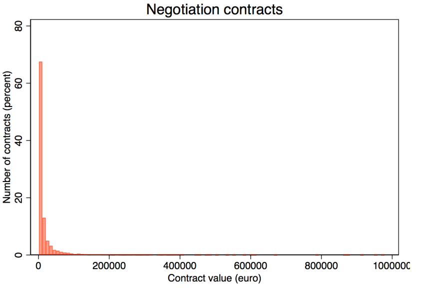


(d) Distribution of value of tenders above 1 mil euro

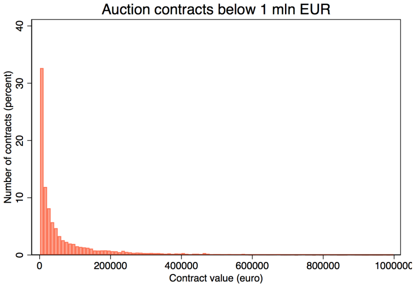
Figure 4.5: Distribution of tender contract values: by tender type(continuation)



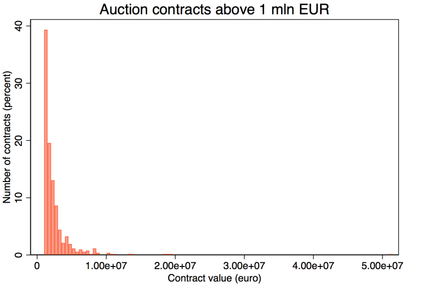
(a) Distribution of value of invitation contacts



(b) Distribution of value of negotiation contracts



(c) Distribution of value of auction contracts below 1 mil euro



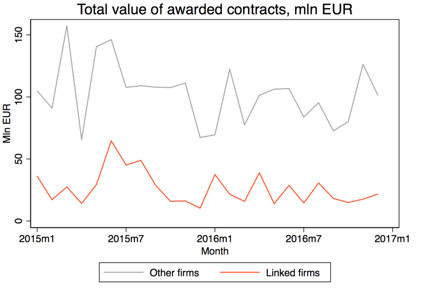
(d) Distribution of value of auction contracts above 1 mil euro

Table 4.14: List of firms with highest value of procurement contracts

Nr.	Firm name	Contracts (mln EUR)	% Total	% Direct
1	**Farmexpert D.C.I.**	383.79	12.4	2.3
2	Mediplus Exim	355.05	11.5	2.2
3	Farmexim	177.66	5.8	5.1
4	Polisano	91.41	3.0	6.6
5	Gadagroup Romania	62.35	2.0	2.5
6	**Fresenius Kabi Romania**	57.35	1.9	3.8
7	Pharmafarm	53.66	1.7	5.9
8	Timi Med	51.41	1.7	0.0
9	**B.Braun Medical**	46.15	1.5	10.5
10	Farmaceutica Remedia Distribution & Logistics	42.43	1.4	3.3
11	**Actavis**	39.79	1.3	2.8
12	Fildas Trading	38.00	1.2	5.9
13	Compania Nationala Unifarm	34.15	1.1	7.4
14	Romastru Trading	33.74	1.1	3.1
15	Silva Trading	31.56	1.0	1.3
16	Sante International	31.37	1.0	3.7
17	Medical Technologies International	31.19	1.0	2.3
18	**Merck Romania**	30.64	1.0	1.7
19	Farmaceutica Remedia	28.36	0.9	2.7
20	Europharm Holding	28.26	0.9	12.4
21	**Roche Romania**	23.89	0.8	2.6
22	Geneva Romfarm International	22.75	0.7	0.6
23	Medical Ortovit	22.44	0.7	4.2
24	Clini-Lab	21.06	0.7	9.8
25	Pharma	20.79	0.7	19.4
26	**Pfizer Romania**	20.26	0.7	1.8
27	Top Diagnostics	20.19	0.7	3.5
28	Cortech Med	19.54	0.6	0.3
29	Siemens	19.34	0.6	1.9
30	One Pharm Grup	18.53	0.6	1.5

Note: bold names indicate the firm has also contracts linked to sponsorships

Figure 4.6: Value of procurement contracts

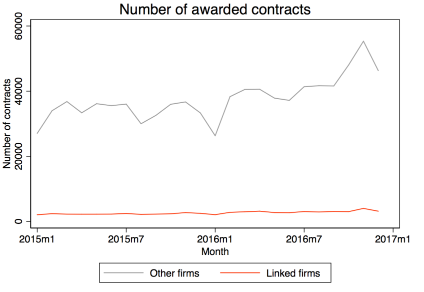


Note: Linked firms are firms with minimum one linked contract in 2015-2016.

Table 4.15: List of product categories with highest value of procurement contracts

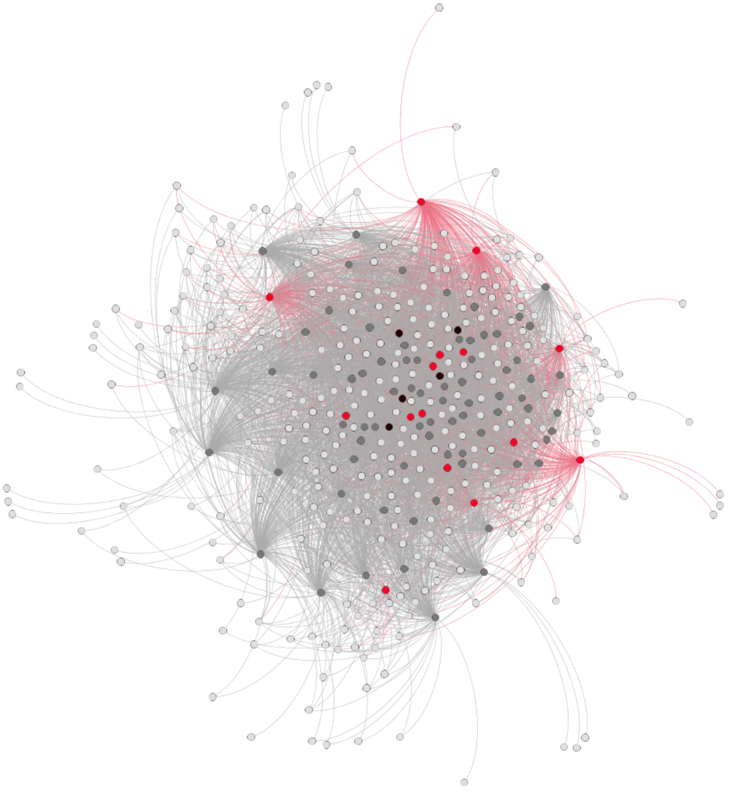
Description	Contracts (mln EUR)
Various Medicinal Products	328.47
Antineoplastic Agents	291.46
Pharmaceutical Products	243.58
Medical Consumables	166.93
Antineoplastic And Immunomodulating Agents	151.94
Laboratory Reagents	109.52
Angioplasty Supplies	90.11
Antivirals For Systemic Use	73.37
Blood-Testing Reagents	70.57
Antibacterials For Systemic Use	68.88

Figure 4.7: Number of procurement contracts



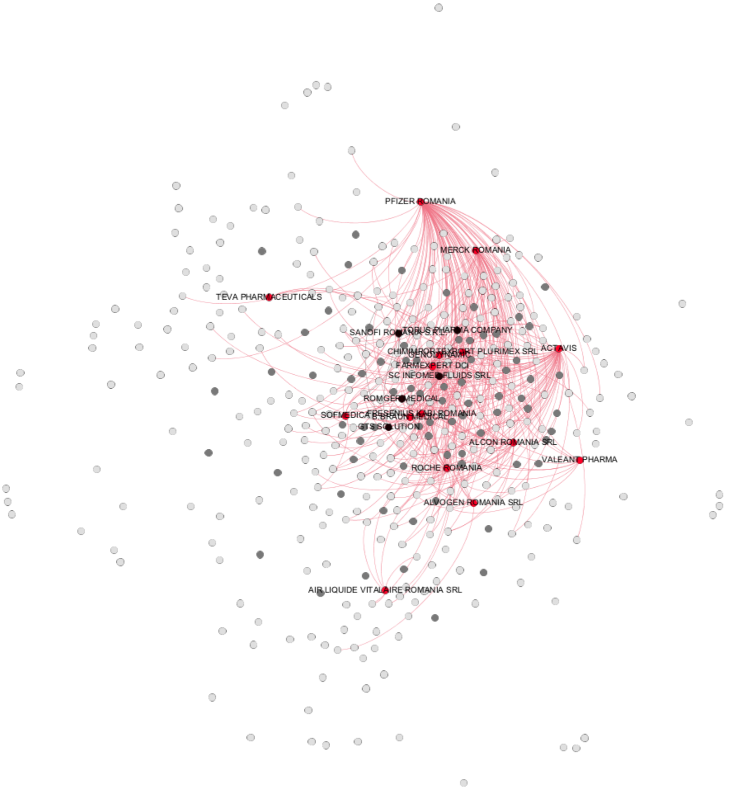
Note: Linked firms are firms with minimum one linked contract in 2015-2016.

Figure 4.8: Network visualisation: all sponsorships



Note: Light grey dots are hospitals, red dots are hospitals with at least one linked contract, black dots are firms that gave sponsorships and received unrelated procurement contracts, dark grey dots are firms that only gave sponsorships. Red lines mean that there is at least one linked contract between the entities, black lines mean that there is a procurement unrelated to the sponsorship, light grey lines are simple sponsorships.

Figure 4.9: Network visualisation: only sponsorships and procurement



Note: Light gray dots are hospitals, red dots are hospitals with at least one linked contract, black dots are firms that gave sponsorships and received unrelated procurement contracts, dark gray dots are firms that only gave sponsorships. Red lines mean that there is at least one linked contract between the entities, black lines mean that there is a procurement unrelated to the sponsorship, light gray lines are simple sponsorships.

4.C Robustness checks

4.C.1 Robustness checks: intensive margin

Table 4.16: The association between sponsorships and contract values: non-linear specification

	(1) lnV	(2) lnV	(3) lnV	(4) lnV
Spons_yes	0.110*			
	(0.058)			
Spons_EUR1000		0.092***	0.122**	0.199**
		(0.028)	(0.055)	(0.092)
Spons_EUR1000_2			-0.004	-0.032
			(0.005)	(0.025)
Spons_EUR1000_3				0.002
				(0.001)
Observations	964952	964952	964952	964952
R^2	0.551	0.551	0.551	0.551
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.17: The association between sponsorships and contract values: including institutional sponsorships

	(1) lnV	(2) lnV	(3) lnV	(4) lnV
spons_all_yes	0.1346***			
	(0.0503)			
Spons_yes_lead		0.0552		
		(0.1112)		
Spons_yes_other		0.0986		
		(0.0607)		
Spons_yes_inst		0.2309***		
		(0.0878)		
spons_all_EUR1000			0.0109	
			(0.0067)	
Spons_EUR1000_lead				0.1032**
				(0.0505)
Spons_EUR1000_other				0.0883***
				(0.0313)
Spons_EUR1000_inst				0.0051
				(0.0053)
Observations	964952	964952	964952	964952
R^2	0.551	0.551	0.551	0.551
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.18: The association between sponsorships and contract values: non-linear specification, heterogeneous effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	lnV	lnV	lnV	lnV	lnV	lnV	lnV	lnV	lnV	lnV
Spons_yes_lead	0.0594									
	(0.1124)									
Spons_yes_other	0.1001									
	(0.0610)									
Spons_EUR1000_lead		-0.1338	0.0681	0.1041**	0.0074	0.1160	0.1230**			
		(0.2961)	(0.1279)	(0.0508)	(0.2999)	(0.1345)	(0.0546)			
Spons_EUR1000_lead_2		0.1230	0.0084		0.0681	0.0020				
		(0.1436)	(0.0238)		(0.1455)	(0.0245)				
Spons_EUR1000_lead_3		-0.0132			-0.0078					
		(0.0148)			(0.0151)					
Spons_EUR1000_other		0.2498**	0.1291**	0.0900***				0.2530**	0.1366**	0.0933***
		(0.1011)	(0.0616)	(0.0313)				(0.1001)	(0.0621)	(0.0315)
Spons_EUR1000_other_2		-0.0476*	-0.0057					-0.0477*	-0.0064	
		(0.0254)	(0.0063)					(0.0253)	(0.0063)	
Spons_EUR1000_other_3		0.0026**						0.0026**		
		(0.0013)						(0.0013)		
Observations	964952	964952	964952	964952	964952	964952	964952	964952	964952	964952
R^2	0.551	0.551	0.551	0.551	0.551	0.551	0.551	0.551	0.551	0.551
clustvar	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.19: The association between sponsorships and contract values: only direct contracts

	(1) lnV	(2) lnV	(3) lnV	(4) lnV
Spons_yes	0.0614			
	(0.0604)			
Spons_yes_lead		-0.0354		
		(0.1036)		
Spons_yes_other		0.0690		
		(0.0625)		
Spons_EUR1000			0.0886^{***}	
			(0.0340)	
Spons_EUR1000_lead				0.1107^{*}
				(0.0572)
Spons_EUR1000_other				0.0838^{**}
				(0.0389)
Observations	936780	936780	936780	936780
R^2	0.449	0.449	0.449	0.449
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.20: The association between sponsorships and contract values: only tender contracts

	(1) lnV	(2) lnV	(3) lnV	(4) lnV
Spons_yes	0.1698			
	(0.1211)			
Spons_yes_lead		0.0823		
		(0.2096)		
Spons_yes_other		0.1538		
		(0.1390)		
Spons_EUR1000			0.0676	
			(0.0413)	
Spons_EUR1000_lead				0.0494
				(0.1236)
Spons_EUR1000_other				0.0698
				(0.0448)
Observations	27837	27837	27837	27837
R^2	0.506	0.506	0.506	0.506
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

4.C.2 Robustness checks: extensive margin

	(1) Any	(2) Any	(3) Direct	(4) Direct	(5) NotDirect	(6) NotDirect
Spons_EUR1000	0.0138***	0.0231***	0.0079*	0.0112	0.0080***	0.0185***
	(0.0051)	(0.0077)	(0.0044)	(0.0079)	(0.0024)	(0.0032)
Spons_EUR1000_2		-0.0011*		-0.0004		-0.0013***
		(0.0006)		(0.0006)		(0.0003)
Observations	1203096	1203096	1203096	1203096	1203096	1203096
R^2	0.130	0.130	0.131	0.131	0.064	0.064
clustvar	spit_id	spit_id	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

4.C.3 Robustness checks: red flags

Table 4.21: Are sponsorships associated with less offers for auctions?

	(1) AvgNrOffers	(2) AvgNrOffers	(3) AvgNrOffers	(4) AvgNrOffers
ContractValue1000	-0.0001 (0.0002)	-0.0001 (0.0002)	-0.0001 (0.0002)	-0.0001 (0.0002)
Spons_yes	-0.7927 (0.4979)			
Spons_yes_lead		-1.1444 (0.8791)		
Spons_yes_other		-0.6848 (0.5534)		
Spons_EUR1000			-0.2379 (0.2034)	
Spons_EUR1000_lead				-0.6411 (0.8608)
Spons_EUR1000_other				-0.1803 (0.1738)
Observations	15439	15439	15439	15439
R^2	0.623	0.623	0.623	0.623
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, firm

Table 4.22: Are sponsorships associated with single-offer tenders?

	(1) SingleOffer	(2) SingleOffer	(3) SingleOffer	(4) SingleOffer
ContractValue1000	0.0000**	0.0000**	0.0000**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spons_yes	0.0210*			
	(0.0126)			
Spons_yes_lead		0.0220		
		(0.0172)		
Spons_yes_other		0.0136		
		(0.0130)		
Spons_EUR1000			-0.0008	
			(0.0037)	
Spons_EUR1000_lead				0.0056
				(0.0071)
Spons_EUR1000_other				-0.0017
				(0.0040)
Observations	21368	21368	21368	21368
R^2	0.383	0.383	0.383	0.383
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.23: Are sponsorships associated with less offers for tenders?

	(1) AvgNrOffers	(2) AvgNrOffers	(3) AvgNrOffers	(4) AvgNrOffers
ContractValue1000	0.0001	0.0001	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Spons_yes	-0.4712			
	(0.4280)			
Spons_yes_lead		-0.8492		
		(0.8541)		
Spons_yes_other		-0.3003		
		(0.4605)		
Spons_EUR1000			-0.1482	
			(0.2043)	
Spons_EUR1000_lead				-0.2969
				(0.9136)
Spons_EUR1000_other				-0.1272
				(0.1740)
Observations	21368	21368	21368	21368
R^2	0.562	0.562	0.562	0.562
clustvar	spit_id	spit_id	spit_id	spit_id

All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Table 4.24: Are sponsorships associated with shorter tenders?

	(1) ProcLength	(2) ProcLength	(3) ProcLength	(4) ProcLength
ContractValue1000	-0.0002	-0.0002	-0.0002	-0.0002
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Spons_yes	-1.9866*			
	(1.1370)			
Spons_yes_lead		-4.4282**		
		(1.9897)		
Spons_yes_other		-1.0720		
		(1.1832)		
Spons_EUR1000			-0.7103	
			(0.4486)	
Spons_EUR1000_lead				-3.1462***
				(1.1466)
Spons_EUR1000_other				-0.3664
				(0.4324)
Observations	21334	21334	21334	21334
R^2	0.695	0.695	0.695	0.695
clustvar	spit_id	spit_id	spit_id	spit_id

ProcLength = absolute difference between Announcement Date and Contract Date. All regressions include the following fixed effects: month, year, hospital, product, procedure, firm

Nederlandse Samenvatting (Summary in Dutch)

Dit proefschrift bestaat uit twee delen: hoofdstukken 2 en 3 bestuderen hoe de overheid belasting dient te heffen om op een doelmatige manier opbrengsten te genereren, terwijl hoofdstuk 4 analyseert hoe publiek geld daadwerkelijk besteed wordt in een sector waar overheidsingrijpen een belangrijke rol speelt: de gezondheidszorg. Deze vragen zijn twee zijden van dezelfde medaille, namelijk het begrijpen hoe overheidsbestedingen doelmatiger te maken. Hoe de overheid geld besteedt is net zo belangrijk als hoe zij geld verwerft. Enerzijds kan, zelfs als belastingen zijn geoptimaliseerd, publiek geld worden verspild door corruptie of ander inefficiënt beleid. Anderzijds kan ondoelmatige belastingheffing leiden tot een tekort aan publieke middelen, wat de overheid beperkt in haar mogelijkheid om essentiële diensten te verlenen.

Hoofdstuk 2 beoogt een intuïtieve verklaring te geven voor het Chamley-Judd resultaat, een bekend theoretisch resultaat dat stelt dat inkomsten uit kapitaal in het lange termijn evenwicht (in de ‘steady state) niet moet worden belast (Chamley, 1986; Judd, 1985). De verklaring volgt uit Corlett and Hague (1953): als goederen die sterk complementair zijn aan vrije tijd zwaarder worden belast, besluiten individuen om minder vrije tijd te genieten en meer te werken. In de steady state zijn huidige en toekomstige consumptie even complementair aan arbeid, onafhankelijk van welke nutsfunctie wordt verondersteld. Hieruit volgt dat een gedifferentieerde belasting op consumptie (ofwel een kapitaalinkomstenbelasting) geen baten genereert bovenop een belasting op arbeidsinkomsten, maar wel leidt tot verstoringen in consumptiebeslissingen.

Hoofdstuk 3 beoogt de modellen van optimale belastingheffing dichterbij de werkelijkheid te brengen. Het hoofdstuk onderzoekt hoe inkomen uit kapitaal worden belast zou moeten worden als individuen verschillende rendementen behalen op hun kapitaal. Door overvloedig empirisch bewijs dat grote verschillen in behaalde rendementen langs de vermogensverdeling laat zien (Fagereng et al., 2016; Campbell, Ramadorai, and Ranish, 2018), wordt deze vraag steeds beleidsrelevanter. Dit hoofdstuk laat zowel theoretisch als numeriek zien dat in deze omstandigheid de optimale belasting op kapitaalinkomen positief en economisch significant is.

Hoofdstuk 4 vormt het tweede deel van dit proefschrift en onderzoekt mogelijke gevallen van beïnvloeding en belangenverstrengeling in Roemeense publieke ziekenhuizen. In het bijzonder kijkt dit hoofdstuk naar de link tussen farmaceutische bedrijven die doktoren in publieke ziekenhuizen sponsoren en de aanbestedingen die deze ziekenhuizen toekennen aan diverse farmaceutische bedrijven. Het sponsoren van een dokter met managementtaken is geassocieerd met een hogere kans op een direct contract (zonder aanbesteding) dan het sponsoren van een reguliere dokter, maar dit verschil is niet economisch significant voor contracten met aanbesteding. Het hoofdstuk documenteert ook een timingseffect: binnen drie maanden na een sponsoring is er een verhoogde kans dat er een contract wordt getekend tussen het gesponsorde ziekenhuis en het bedrijf dat sponsort. Bovendien zijn met contracten die worden gelinkt aan een sponsoring grotere bedragen gemoeid dan met contracten die niet zijn gelinkt. Gezien de institutionele setting en verder bewijs dat suggereert dat contracten die zijn gelinkt aan een sponsoring minder transparant zijn, lijkt het erop dat sponsoring meer dient als een verzoek om een wederdienst dan als legitiem marketingmiddel.

Bibliography

Abel, Andrew B. 2007. "Optimal Capital Income Taxation." Working Paper 13354, National Bureau of Economic Research, Cambridge, MA.

Abraham, John. 2002. "The Pharmaceutical Industry as a Political Player." The Lancet 360 (9344):1498–1502.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson. 2009. "The Age of Reason: Financial Decisions Over the Life Cycle and Implications for Regulation." Brookings Papers on Economic Activity 2009 (2):51–117.

Aiyagari, S. Rao. 1995. "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints and Constant Discounting." Journal of Political Economy 103 (6):1158–1175.

Albanesi, Stefania and Roc Armenter. 2009. "Intertemporal Distortions in the Second Best." Working Paper 13629, National Bureau of Economic Research, Cambridge, MA.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. World Inequality Report 2018. World Inequality Lab.

Arfken, George B. and Hans J. Weber. 2005. Mathematical Methods for Physicists International Student Edition. Elsevier, Burlington, MA.

Asher, Sam and Paul Novosad. 2017. "Politics and Local Economic Growth: Evidence from India." American Economic Journal: Applied Economics 9 (1):229–73.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." Journal of Economic Literature 49 (1):3–71.

Atkinson, Anthony B. and Agnar Sandmo. 1980. "Welfare Implications of the Taxation of Savings." Economic Journal 90 (359):529–549.

Atkinson, Anthony B. and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct Versus Indirect Taxation." Journal of Public Economics 6 (1-2):55–75.

———. 1980. Lectures on Public Economics. London and New York: McGraw-Hill.

Attanasio, Orazio P. and Guilermo Weber. 2010. "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy." Journal of Economic Literature 48 (3):693–751.

Auerbach, Alan J. and Laurence J. Kotlikoff. 1983. "National Savings, Economic Welfare, and the Structure of Taxation." In Behavioral Simulation Methods in Tax Policy Analysis, edited by Martin Feldstein, NBER Chapters, chap. 13. Cambridge, MA.: University of Chicago Press, 459–498.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2018. "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." Research Paper FIN-2016-1126, HEC Paris, Paris, France.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." American Economic Review 99 (4):1278–1308.

Banks, James and Peter A. Diamond. 2010. "The Base for Direct Taxation." In The Mirrlees Review. Dimensions of Tax Design, edited by James A. Mirrlees et al., chap. 6. Oxford: Oxford University Press, 548–648.

Barber, Brad M., Terrance Odean, and Lu Zheng. 2005. "Out of Sight, Out of Mind: the Effects of Expenses on Mutual Fund Flows." Journal of Business 78 (6):2095–2120.

Benartzi, Shlomo and Richard H. Thaler. 2001. "Naive Diversification Strategies in Defined Contribution Saving Plans." American Economic Review 91 (1):79–98.

Benhabib, Jess, Alberto Bisin, and Mi Luo. 2019. "Wealth Distribution and Social Mobility in the US: A Quantitative Approach." American Economic Review 109 (5):1623–47.

Blomquist, Sören and Håkan Selin. 2010. "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates." Journal of Public Economics 94 (11):878–889.

Blundell, Richard and Thomas MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In Handbook of Labor Economics, vol. 3, edited by Orley C. Ashenfelter and David Card. Amsterdam, the Netherlands: Elsevier, 1559–1695.

Bohácek, Radim and Michal Kejak. 2016. "Optimal Government Policies in Models with Heterogeneous Agents." Working Paper 272, CERGE-EI, Prague.

Bruelhart, Marius, Jonathan Gruber, Matthias Krapf, and Kurt Schmidheiny. 2016. "Taxing Wealth: Evidence from Switzerland." Working Paper 22376, National Bureau of Economic Research, Cambridge, MA.

Burguet, Roberto and Yeon-Koo Che. 2004. "Competitive Procurement with Corruption." The RAND Journal of Economics 35 (1):50–68.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." Journal of Political Economy 115 (5):707–747.

Campbell, John Y. 2016. "Restoring Rational Choice: the Challenge of Consumer Financial Regulation." American Economic Review: Papers & Proceedings 106 (5):1–30.

Campbell, John Y., Tarun Ramadorai, and Benjamin Ranish. 2018. "Do the Rich Get Richer in the Stock Market? Evidence from India." Working Paper 24898, National Bureau of Economic Research, Cambridge, MA.

Castro, Massimo Finocchiaro, Calogero Guccio, and Ilde Rizzo. 2014. "An Assessment of the Waste Effects of Corruption on Infrastructure Provision." International Tax and Public Finance 21 (4):813–843.

Chamley, Christophe. 1986. "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." Econometrica 54 (3):607–622.

Chari, Varadarajan V. and Patrick J. Kehoe. 1999. "Optimal Fiscal and Monetary Policy." In Handbook of Macroeconomics, Handbook of Macroeconomics, vol. 1, edited by John B. Taylor and Michael Woodford, chap. 26. Amsterdam: Elsevier, 1671–1745.

Chari, Varadarajan V., Juan P. Nicolini, and Pedro Teles. 2018. "Optimal Capital Taxation Revisited." Staff Report 571, Federal Reserve Bank of Minneapolis, Minneapolis, MN.

Choi, James J., David Laibson, and Brigitte C. Madrian. 2010. "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds." Review of Financial Studies 23 (4):1405–1432.

———. 2011. "$100 Bills On the Sidewalk: Suboptimal Investment in 401 (k) Plans." Review of Economics and Statistics 93 (3):748–763.

Choi, James J., David I. Laibson, and Brigitte C. Madrian. 2005. "Are Empowerment and Education Enough? Underdiversification in 401 (k) Plans." Brookings Papers on Economic Activity 2005 (2):151–213.

Christiansen, Vidar. 1981. "Evaluation of Public Projects under Optimal Taxation." Review of Economic Studies 48 (3):447–457.

———. 1984. "Which Commodity Taxes Should Supplement the Income Tax?" Journal of Public Economics 24 (2):195–220.

———. 1993. "A Normative Analysis of Capital Income Taxes in the Presence of Aggregate Risk." Geneva Papers on Risk and Insurance Theory 18 (1):55–76.

Christiansen, Vidar and Matti Tuomala. 2008. "On Taxing Capital Income with Income Shifting." International Tax and Public Finance 15 (4):527–545.

Cole, Shawn and Anh Tran. 2011. "Evidence from the Firm: A New Approach to Understanding Corruption." International Handbook on the Economics of Corruption :408.

Conesa, Juan C., Sagiri Kitao, and Dirk Krueger. 2009. "Taxing Capital? Not a Bad Idea After All!" American Economic Review 99 (1):25–48.

Corlett, Wilfred J. and Douglas C. Hague. 1953. "Complementarity and the Excess Burden of Taxation." Review of Economic Studies 21 (1):21–30.

Correia, Isabel H. 1996. "Should Capital Income be Taxed in the Steady State?" Journal of Public Economics 60 (1):147–151.

Cotlear, Daniel, Somil Nagpal, Owen Smith, Ajay Tandon, and Rafael Cortez. 2015. Going Universal: How 24 Developing Countries Are Implementing Universal Health Coverage from the Bottom Up. Washington, DC.: The World Bank.

Coviello, Decio and Mario Mariniello. 2014. "Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design." Journal of Public Economics 109:76–100.

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet. 2001. "Direct versus Indirect Taxation: The Design of Tax Structure Revisited." International Economic Review 42 (3):781–799.

Crossley, Thomas F. and Hamish W. Low. 2011. "Is the Elasticity of Intertemporal Substitution Constant?" Journal of the European Economic Association 9 (1):87–105.

Deaton, Angus. 1979. "Optimally Uniform Commodity Taxes." Economics Letters 2 (4):357–361.

DeJong, Colette, Thomas Aguilar, Chien-Wen Tseng, Grace A Lin, W John Boscardin, and R Adams Dudley. 2016. "Pharmaceutical Industry–sponsored Meals and Physician Prescribing Patterns for Medicare Beneficiaries." JAMA Internal Medicine 176 (8):1114–1122.

DeNavas-Walt, Carmen and Bernadette D. Proctor. 2015. “Income and Poverty in the United States: 2014.” Current Population Reports, U.S. Census Bureau, Suitland, MD.

Desai, Raj M and Anders Olofsgård. 2011. “The Costs of Political Influence: Firm-level Evidence from Developing Countries.” Quarterly Journal of Political Science 6 (2):137–178.

Diamond, Peter A. 1998. “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates.” American Economic Review 88 (1):83–95.

Diamond, Peter A. and James A. Mirrlees. 1971a. “Optimal Taxation and Public Production I: Production Efficiency.” American Economic Review 61 (1):8–27.

———. 1971b. “Optimal Taxation and Public Production II: Tax Rules.” American Economic Review 61 (3):261–78.

———. 1978. “A Model of Social Insurance with Variable Retirement.” Journal of Public Economics 10 (3):295–336.

Diamond, Peter A. and Emmanuel Saez. 2011. “The Case for a Progressive Tax: From Basic Research to Policy Recommendations.” Journal of Economic Perspectives 25 (4):165–190.

Diamond, Peter A. and Johannes Spinnewijn. 2011. “Capital Income Taxes with Heterogeneous Discount Rates.” American Economic Journal: Economic Policy 3 (4):52–76.

Dieleman, Joseph L., Tara Templin, Nafis Sadat, Patrick Reidy, Abigail Chapin, Kyle Foreman, Annie Haakenstad, Tim Evans, Christopher JL Murray, and Christoph Kurowski. 2016. “National Spending on Health by Source for 184 Countries Between 2013 And 2040.” The Lancet 387 (10037):2521–2535.

DNA. 2017a. “Abuzul In Serviciu In Domeniul Achizitiilor Publice.” Tech. rep., Directia Nationala Anticoruptie.

———. 2017b. “Press Release 128/VIII/3.”

Dowd, Tim, Robert McClelland, and Athiphat Muthitacharoen. 2012. “New Evidence on the Elasticity of Capital Gains.” Working Paper 2012-09, Congressional Budget Office, Washington, DC.

Egan, Mark, Gregor Matvos, and Amit Seru. 2018. “The Market for Financial Adviser Misconduct.” Journal of Political Economy 127 (1):233–295.

Erosa, Andrès and Martin Gervais. 2002. “Optimal Taxation in Life-Cycle Economies.” Journal of Economic Theory 105 (2):338–369.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. “Heterogeneity and Persistence in Returns to Wealth.” Working Paper 22822, National Bureau of Economic Research, Cambridge, MA. Mimeo.

———. 2020. “Heterogeneity and Persistence in Returns to Wealth.” Econometrica (Forthcoming).

Fazekas, Mihaly and Istvan Janos Toth. 2016. “From Corruption to State Capture: A New Analytical Framework with Empirical Applications from Hungary.” Political Research Quarterly 69 (2):320–334.

Federal Reserve Bank of St. Louis. 2020. “30-Year Treasury Constant Maturity Rate, Percent, Annual, Not Seasonally Adjusted.” Federal Reserve Economic Data, Federal Reserve Bank of St. Louis.

Ferraz, Claudio and Frederico Finan. 2008. “Exposing Corrupt Politicians: The Effects of Brazil’s Publicly Released Audits on Electoral Outcomes.” Quarterly Journal of Economics 123 (2):703–745.

Fischer, Stanley. 1980. “Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government.” Journal of Economic Dynamics and Control 2:93–107.

Fisman, Raymond and Roberta Gatti. 2002. “Decentralization and Corruption: Evidence Across Countries.” Journal of Public Economics 83 (3):325–345.

Fisman, Raymond and Edward Miguel. 2007. “Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets.” Journal of Political economy 115 (6):1020–1048.

Flood, Lennart and Thomas MaCurdy. 1992. “Work Disincentive Effects of Taxes: An Empirical Analysis of Swedish Men.” Carnegie-Rochester Conference Series on Public Policy 37:239–277.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. “The Dynamics of Inequality.” Econometrica 84 (6):2071–2111.

Gahvari, Firouz and Luca Micheletto. 2016. “Capital Income Taxation and the Atkinson–Stiglitz Theorem.” Economics Letters 147:86–89.

Gerritsen, Aart. 2016. “Optimal Taxation when People Do Not Maximize Well-being.” Journal of Public Economics 144:122–139.

Gerritsen, Aart, Bas Jacobs, Alexandra Rusu, and Kevin Spiritus. 2019. “Optimal Taxation of Capital Income with Heterogeneous Rates of Return.” Mimeo, Erasmus University Rotterdam, Rotterdam, the Netherlands.

Goetzmann, William N. and Alok Kumar. 2008. "Equity Portfolio Diversification." Review of Finance 12 (3):433–463.

Goldberg, Felix. 2017. "Corruption and Lobbying: Conceptual Differentiation and Gray Areas." Crime, Law and Social Change 70 (2):197–215.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." Review of Economic Studies 70 (3):569–587.

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin. 2014. "A Variational Approach to the Analysis of Tax Systems." Working Paper 20780, National Bureau of Economic Research, Cambridge, MA.

Gordon, Roger H. 1985. "Taxation of Corporate Capital Income: Tax Revenues Versus Tax Distortions." Quarterly Journal of Economics 100 (1):1–27.

Gorman, W.M. 1961. "On a Class of Preference Fields." Metroeconomica 13 (2):53–56.

Gourinchas, Pierre-Olivier and Jonathan A. Parker. 2002. "Consumption Over the Life Cycle." Econometrica 70 (1):47–89.

Gruber, Jonathan. 2006. "A Tax-Based Estimate of the Elasticity of Intertemporal Substitution." Working Paper 11945, National Bureau of Economic Research, Cambridge, MA.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen. 2019. "Use It or Lose It: Efficiency Gains from Wealth Taxation." Working Paper 26284, National Bureau of Economic Research, Cambridge, MA.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2017. "Optimal Tax Progressivity: An Analytical Framework." Quarterly Journal of Economics 132 (4):1693–1754.

Hubbard, R. Glenn and Kenneth L. Judd. 1986. "Liquidity Constraints, Fiscal Policy, and Consumption." Brookings Papers on Economic Activity 1986 (1):1–59.

Immervoll, Herwig, Henrik J. Kleven, Claus T. Kreiner, and Emmanuel Saez. 2007. "Welfare Reform in European Countries: A Microsimulation Analysis." Economic Journal 117 (516):1–44.

Jacobs, Bas. 2013. "The Marginal Cost of Public Funds is One at the Optimal Tax System." Mimeo, Erasmus University Rotterdam, Rotterdam, the Netherlands.

Jacobs, Bas and Robin Boadway. 2014. "Optimal Linear Commodity Taxation under Optimal Non-Linear Income Taxation." Journal of Public Economics 117 (1):201–210.

Jacobs, Bas and Lans A. Bovenberg. 2010. "Human Capital and Optimal Positive Taxation of Capital Income." International Tax and Public Finance 17 (5):451–478.

Jacobs, Bas and Alexandra V. Rusu. 2018. "Why is the Long-Run Tax on Capital Income Zero? Explaining the Chamley-Judd Result." Mimeo, Erasmus University Rotterdam.

Jacobs, Bas and Dirk Schindler. 2012. "On the Desirability of Taxing Capital Income in Optimal Social Insurance." Journal of Public Economics 96 (9–10):853–868.

Jacquet, Laurence and Etienne Lehmann. 2017. "Optimal Income Taxation with Composition Effects." Working Paper 6654, CESifo, Munich, Germany.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden. 2013. "Optimal Redistributive Taxation with Both Intensive and Extensive Responses." Journal of Economic Theory 148 (5):1770–1805.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. 2018. "Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark." Working Paper 24371, National Bureau of Economic Research, Cambridge, MA.

Jaumotte, Florence. 2003. "Female Labour Force Participation." Working Paper 376, OECD, Paris.

Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi. 1997. "On the Optimal Taxation of Capital Income." Journal of Economic Theory 73 (1):93–117.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." Journal of Public Economics 28 (1):59–83.

———. 1999. "Optimal Taxation and Spending in General Competitive Growth Models." Journal of Public Economics 71 (3):1–26.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens. 2018. "Investor Sophistication and Capital Income Inequality." Journal of Monetary Economics 107:18–31.

King, Mervyn. A. 1980. "Savings and Taxation." In Public Policy and the Tax System, edited by Gordon A. Hughes and Geoffrey M. Heal. London: Allen and Unwin, 1–35.

Kristjánsson, Arnaldur S. 2016. "Optimal Taxation with Endogenous Return to Capital." Mimeo, Oslo University.

Krusell, Per, Burhanettin Kuruşçu, and Anthony A. Smith. 2010. "Temptation and Taxation." Econometrica 78 (6):2063–2084.

Kydland, Finn E. and Edward C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." Journal of Political Economy 85 (3):473–492.

Lansing, Kevin J. 1999. "Optimal Redistributive Capital Taxation in a Neoclassical Growth Model." Journal of Public Economics 73 (3):423–453.

Larkin, Ian, Desmond Ang, Jonathan Steinhart, Matthew Chao, Mark Patterson, Sunita Sah, Tina Wu, Michael Schoenbaum, David Hutchins, Troyen Brennan et al. 2017. "Association Between Academic Medical Center Pharmaceutical Detailing Policies and Physician Prescribing." JAMA 317 (17):1785–1795.

Lehmann, Etienne, Sander Renes, Kevin Spiritus, and Floris T. Zoutman. 2018. "Optimal Joint Taxation." Mimeo, Erasmus University Rotterdam.

Lehne, Jonathan, Jacob N. Shapiro, and Oliver Van den Eynde. 2018. "Building Connections: Political Corruption and Road Construction in India." Journal of Development Economics 131:62–78.

Ljungqvist, Lars and Thomas J. Sargent. 2004. Recursive Macroeconomic Theory. Cambridge, MA.: MIT-Press.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S. Mitchell. 2017. "Optimal Financial Knowledge and Wealth Inequality." Journal of Political Economy 125 (2):431–477.

Lusardi, Annamaria and Olivia S. Mitchell. 2011. "Financial Literacy Around the World: An Overview." Journal of Pension Economics and Finance 10 (4):497–508.

Mankiw, N. Gregory, Matthew C. Weinzierl, and Danny Yagan. 2011. "Optimal Taxation in Theory and Practice." Journal of Economic Perspectives 23 (4):147–174.

Meghir, Costas and David Phillips. 2010. "Labour Supply and Taxes." In The Mirrlees Review. Dimensions of Tax Design, edited by James A. Mirrlees et al. Oxford, United Kingdom: Oxford University Press, 202–274.

Mill, John S. 1848. Principles of Political Economy. London, United Kingdom: John W. Parker.

Mironov, Maxim and Ekaterina Zhuravskaya. 2016. "Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data." American Economic Journal: Economic Policy 8 (2):287–321.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." Review of Economic Studies 38 (2):175–208.

———. 1976. “Optimal Tax Theory: A Synthesis.” Journal of Public Economics 6 (4):327–358.

Mirrlees, James A., Stuart Adam, Timothy J. Besley, Richard Blundell, Steven Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth D. Myles, and James M. Poterba. 2011. Tax by Design: The Mirrlees Review. Oxford: Oxford University Press.

Mulligan, Casey B. 2002. “Capital, Interest, and Aggregate Intertemporal Substitution.” Working Paper 9373, National Bureau of Economic Research, Cambridge, MA.

Olken, Benjamin A. 2006. “Corruption and the Costs of Redistribution: Micro Evidence from Indonesia.” Journal of Public Economics 90 (45):853–870.

Ordover, Janusz A. and Edmund S. Phelps. 1979. “The Concept of Optimal Taxation in the Overlapping–Generations Model of Capital and Wealth.” Journal of Public Economics 12 (1):1–26.

Palguta, Ján and Filip Pertold. 2017. “Manipulation of Procurement Contracts: Evidence from the Introduction of Discretionary Thresholds.” American Economic Journal: Economic Policy 9 (2):293–315.

Pigou, Arthur C. 1928. “An Analysis of Supply.” Economic Journal 38 (150):238–257.

Piketty, Thomas. 2014. Capital in the Twenty-First Century. Cambridge, MA.: Harvard University Press.

Piketty, Thomas and Gabriel Zucman. 2013. “Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010.” Discussion Paper 9588, Centre for Economic Policy Research.

Ramsey, Frank P. 1927. “A Contribution to the Theory of Taxation.” Economic Journal 37 (145):47–61.

Reinikka, Ritva and Jakob Svensson. 2004. “Local Capture: Evidence from a Central Government Transfer Program in Uganda.” Quarterly Journal of Economics 119 (2):679–705.

———. 2006. “Using Micro-Surveys to Measure and Explain Corruption.” World Development 34 (2):359–370. Part Special Issue (pp. 324404). Corruption and Development: Analysis and Measurement.

Reis, Catarina. 2010. “Entrepreneurial Labor and Capital Taxation.” Macroeconomic Dynamics 15 (3):326–335.

Renes, Sander and Floris T. Zoutman. 2014. “When a Price is Enough: Implementation in Optimal Tax Design.” Discussion Paper TI 14-121/VII, Tinbergen Institute, Amsterdam, the Netherlands.

RISE Project. 2018. "Institutul de Boli Inventate." Online.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." Review of Economic Studies 68 (1):205–229.

———. 2002. "The Desirability of Commodity Taxation Under Non-linear Income Taxation and Heterogeneous Tastes." Journal of Public Economics 83 (2):217–230.

Saez, Emmanuel and Stefanie Stantcheva. 2018. "A Simpler Theory of Optimal Capital Taxation." Journal of Public Economics 162:120–142.

Saez, Emmanuel and Gabriel Zucman. 2016. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." Quarterly Journal of Economics 131 (2):519–578.

Sandmo, Agnar. 1974. "A Note on the Structure of Optimal Taxation." American Economic Review 64 (4):701–706.

Scheuer, Florian and Iván Werning. 2017. "The Taxation of Superstars." Quarterly Journal of Economics 132 (1):211–270.

Schindler, Dirk. 2008. "Taxing Risky Capital Income - A Commodity Taxation Approach." FinanzArchiv 64 (3):311–333.

Seim, David. 2017. "Behavioral Responses to Wealth Taxes: Evidence from Sweden." American Economic Journal: Economic Policy 9 (4):395–421.

Spiritus, Kevin and Robin Boadway. 2017. "The Optimal Taxation of Risky Capital Income: The Rate of Return Allowance." Working Paper 6297, CESifo, Munich, Germany.

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." Journal of Public Economics 17 (2):213–240.

———. 1985. "Inequality and Capital Taxation." Technical Report 457, Stanford University Institute for Mathematical Studies in the Social Sciences.

———. 2000. Economics of the Public Sector – Third Edition. New York/London: W.W. Norton & Company.

———. 2018. "Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution." Journal of Public Economics 162:101–119.

Straub, Ludwig and Iván Werning. 2020. "Positive Long Run Capital Taxation: Chamley-Judd Revisited." American Economic Review 110 (1):86–119.

Tabakovic, Haris and Thomas G. Wollmann. 2018. "From Revolving Doors to Regulatory Capture? Evidence from Patent Examiners." Working Paper 24638, National Bureau of Economic Research, Cambridge, MA.

Transparency International. 2017. "Corruption Perceptions Index." Tech. rep., Transparency International.

U.S. Census Bureau. 2015. "Current Population Survey." 2015 annual social and economic supplement, U.S. Census Bureau, Suitland, MD.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." Journal of Financial Economics 101 (2):449–472.

Varian, Hal R. 1980. "Redistributive Taxation as Social Insurance." Journal of Public Economics 14 (1):49–68.

Vermeulen, Philip. 2018. "How Fat is the Top Tail of the Wealth Distribution?" Review of Income and Wealth 64 (2):357–387.

Von Gaudecker, Hans-Martin. 2015. "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?" Journal of Finance 70 (2):489–507.

Werning, Ivàn. 2007. "Optimal Fiscal Policy with Redistribution." Quarterly Journal of Economics 122 (3):925–967.

Yitzhaki, Shlomo. 1987. "The Relation Between Return and Income." Quarterly Journal of Economics 102 (1):77–95.

Zoutman, Floris T. 2018. "The Effect of Capital Income Taxation on Household Savings." Mimeo, Norwegian School of Economics.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

710. N. GHEBRIHIWET, *Multinational Firms, Technology Transfer, and FDI Policy*
711. H.FANG, *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*
712. Y. KANTOR, *Urban Form and the Labor Market*
713. R.M. TEULINGS, *Untangling Gravity*
714. K.J.VAN WILGENBURG, *Beliefs, Preferences and Health Insurance Behavior*
715. L. SWART, *Less Now or More Later? Essays on the Measurement of Time Preferences in Economic Experiments*
716. D. NIBBERING, *The Gains from Dimensionality*
717. V. HOORNWEG, *A Tradeoff in Econometrics*
718. S. KUCINSKAS, *Essays in Financial Economics*
719. O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*
720. E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*
721. M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*
722. N. CIURILA, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*
723. N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*
724. S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*
725. S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*
726. Y.ZHU, *On the Effects of CEO Compensation*

727. S. XIA, *Essays on Markets for CEOs and Financial Analysts*

728. I. SAKALAUSKAITE, *Essays on Malpractice in Finance*

729. M.M. GARDBERG, *Financial Integration and Global Imbalances*

730. U. THUMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change,*

731. B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*

732. G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*

733. Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*

734. C.M. OOSTERVEEN, *Education Design Matters*

735. S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*

736. S. SOVAGO, *Where to Go Next? Essays on the Economics of School Choice*

737. M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*

738. M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*

739. R.J. DOTTLING, *Essays in Financial Economics*

740. E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*

741. Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*

742. A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*

743. B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*

744. R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*

745. P.M. GOLEC, *Essays in Financial Economics*

746. M.N. SOUVERIJN, *Incentives at work*

747. M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*

748. P. VAN BRUGGEN, *On Measuring Preferences*

749. M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*

750. S. GARCIA MANDICO, *Social Insurance, Labor Supply and Intra-Household Spillovers*

751. Y. SUN, Consumer Search and Quality

752. I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*

753. G.W. GOY, *Modern Challenges to Monetary Policy*

754. A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*

755. J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*

756. J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*

757. A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*

758. F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*

759. A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*

The first part of this thesis covers a question currently high on the public agenda: whether and how to tax capital income. By reinterpreting the Chamley-Judd result, a well-known result in public finance which argues against taxing capital income, it shows that the steady-state assumption is more important than previously thought. Then, it studies how capital income should be taxed when returns to capital differ across individuals, for instance because capital income is positively correlated with ability, or because of returns to scale in investment. Using numerical simulations and economic theory, it concludes that the optimal tax rate on capital income is positive and economically significant. The second part of the thesis studies how public funds are actually spent, investigating possible instances of conflict of interest in the pharmaceutical procurement market. It documents a timing effect between sponsorships offered by pharmaceutical companies to doctors in public hospitals and the procurement contracts received by the companies.

Alexandra Rusu (1989) wrote her Ph.D. dissertation under the supervision of Prof. Bas Jacobs and Prof. Dinand Webbink, at Erasmus University Rotterdam and the Tinbergen Institute. She holds an M.Phil. degree in economics from the Tinbergen Institute and a B.Sc. in economics and business economics from Utrecht University.