Federal Reserve Bank of Chicago

A Fair Day's Pay for a Fair Day's Work: Optimal Tax Design as Redistributional Arbitrage

Christian Hellwig and Nicolas Werquin

January 6, 2022

WP 2022-03

https://doi.org/10.21033/wp-2022-03

A Fair Day's Pay for a Fair Day's Work: Optimal Tax Design as Redistributional Arbitrage*

Christian Hellwig

Toulouse School of Economics

Nicolas Werquin

Federal Reserve Bank of Chicago, Toulouse School of Economics, and CEPR

January 6, 2022

Abstract

We study optimal tax design based on the idea that policy-makers face trade-offs between multiple margins of redistribution. Within a Mirrleesian economy with earnings, consumption and retirement savings, we derive a novel formula for optimal income and savings distortions based on redistributional arbitrage. We establish a sufficient statistics representation of the labor income and capital tax rates on top income earners in dynamic environments, which relies on the observed distributions of both income and consumption. Because consumption has a thinner Pareto tail than income, our quantitative results suggest that it is optimal to shift a substantial fraction of the top earners' tax burden from income to savings.

*We thank Mark Aguiar, Alexandre Gaillard, Mike Golosov, Philipp Wangner for useful comments. We are especially grateful to Florian Scheuer. Opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. Christian Hellwig acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010).

"*Our Nation ... should be able to devise ways and means of insuring to all our able-bodied working men and women a fair day's pay for a fair day's work.*"

Franklin D. Roosevelt, Message to Congress on Establishing Minimum Wages and Maximum Hours, 1937

1 Introduction

The importance of both income and leisure for workers' welfare is a central tenet of modern economics. Its empirical validity is well documented, for example, in the centuries-long campaign of the labor movement to improve the welfare of the working classes. Their 19th century slogan "A Fair Day's Pay for a Fair Day's Work" epitomizes this joint concern for wages along with working hours, or leisure, that permeated policy discussions over labor regulation and the concurrent emergence of the welfare state. The slogan was picked up by Roosevelt in a speech that led to the Fair Labor Standards Act (1938), which simultaneously introduced a minimum wage and regulations on total working hours. Contemporary concerns for "work-life balance" suggest that high income earners today value leisure much like their working class peers in the 1930s or the 19th century, and employers acknowledge these concerns when granting workers leisure-related perks or non-pecuniary benefits, work-time flexibility or time-saving benefits like child-care services to working parents.[1]

Aguiar and Hurst (2007) document a large increase in leisure inequality from the top to the bottom of the distribution since the 1960s in the U.S., mirroring the concurrent, well-documented and widely discussed rise in income inequality. This suggests that policy responses to these trends, and optimal policy design more generally, should focus on the redistribution of leisure as much as on the redistribution of income.

Instead, originating with Mirrlees (1971), the problem of optimally designing taxes and social insurance programs is formalized as a tradeoff between the social benefits of redistributing financial

[1]According to Cambridge online dictionary, work-life balance represents *"the amount of time you spend doing your job compared with the amount of time you spend with your family and doing things you enjoy."* A 2011 report by the Council of Economic Advisors (Romer (2011)) reviews evidence suggesting that both employers and employees benefit from improved work-life balance: "*A study of more than 1,500 U.S. workers reported that nearly a third considered work-life balance and flexibility to be the most important factor in considering job offers. In another survey of two hundred human resource managers, two-thirds cited family-supportive policies and flexible hours as the single most important factor in attracting and retaining employees.*" The report itself is evidence that the importance of leisure for employee welfare is recognized at the highest levels of economic policy. The ongoing pandemic provides further evidence of the importance of leisure time for workers' wellbeing: while the time savings and flexibility gains associated with remote work are greeted as a significant improvement in work-life balance, lack of access to child care and home schooling due to school closures are viewed as adding stress to working parents' lives. Schieman et al. (2021) provide evidence from a sample of about 2000 Canadian households that reported work-life balance improved for most workers, excepted for those with children under the age of 12 who reported no change. Their cross-sectional controls further highlight that reported work-life balance appears to be as much affected by working hours and flexibility as it is by financial stress, but unrelated to income after controlling for other job characteristics.

resources from richer to poorer households, and the efficiency costs associated with allocative distortions that such redistribution necessarily entails when these agents' productivity types or inclination to work are not directly observable. The optimal tax schedule equates the marginal efficiency cost of tax distortions at each income level to the marginal benefit of redistribution from higher to lower incomes. One of the most celebrated achievements of this literature has been the derivation of the optimal tax rate on top income earners by Saez (2001) in terms of three observable statistics that give empirical meaning to this tradeoff between efficiency and redistribution: the elasticities of labor supply with respect to marginal and average tax rates (substitution and income effects), and the Pareto coefficient of the tail of the earnings distribution, which measures the degree of top income inequality.

In this paper, we develop an alternative, complementary perspective on optimal tax design, based on the idea that optimal policy design trades off between multiple dimensions of worker welfare. Mirrlees and Saez focused on redistributing after-tax income, or consumption, towards lower income households. However, the planner affects social welfare not only by redistributing consumption, but also by redistributing leisure time from richer towards poorer households. Viewed in this light, the optimal tax schedule equates the marginal costs of the tax distortions to the marginal benefits of transfering leisure from individuals with high earnings to those with low earnings.

A direct corollary of this alternative way of measuring the marginal benefits from redistribution is that the optimal tax system also equates the marginal benefit of redistributing consumption to the marginal benefit of redistributing leisure from higher towards lower types. We call this property of the optimal tax system *redistributional arbitrage*: at the optimum, the marginal welfare gains from additional redistribution must be equalized across different goods, since otherwise the tax designer would have an "arbitrage opportunity" by increasing redistribution along one margin and reducing it along a different one.

Formally, we extend the canonical Mirrleesian tax design problem to allow for two separate consumption goods, which we interpret as "consumption" and "savings". There are two periods. Agents are characterized by a privately observed productivity type. They work, consume and save in period 1, and consume their savings in period 2. The optimal tax design then requires both a non-linear income tax that introduces a wedge between the disutility of labor effort and the marginal utility of consumption, and a non-linear savings tax that introduces a wedge between the marginal utility of consumption and the marginal utility of savings. The model nests the static Mirrleesian tax design problem in the limit where either consumption or savings vanish from the agents' preferences.

As our central result, we derive a novel representation of optimal tax distortions based on redistributional arbitrage. We express the optimal labor distortion τ_Y as

$$1 - \tau_Y = \frac{B_Y}{B_C}$$

where B_Y represents the planner's marginal benefit of redistributing leisure, rescaled by the marginal utility of leisure, and B_C represents the planner's marginal benefit of redistributing consumption, rescaled by the marginal utility of consumption. This is nothing other than saying that the un-rescaled marginal benefits of redistributing consumption and leisure must be equalized.

Along similar lines we represent the optimal savings wedge by equalizing the marginal benefit of redistributing savings to the marginal benefit of redistributing consumption. This representation provides a converse to the uniform commodity taxation principle (Atkinson and Stiglitz (1976)) by identifying how optimal commodity taxation is shaped by departures from their baseline assumptions of homogeneous preferences.

Following Saez (2001), we show that the rescaled marginal benefits can all be mapped to the distribution of allocations, along with key preference elasticities. This allows us to give empirical content to the idea of redistributional arbitrage.

While not explicit in our optimal tax formula, efficiency costs of tax distortions are still very much a part of the analysis since they are incorporated into how the planner evaluates the marginal gains from redistribution: fixing an (incentive-compatible) baseline allocation, the planner cannot freely increase redistribution, but must do so in a manner that preserves incentive compatibility. Following Hellwig (2021), this class of second-best, "incentive-compatibility preserving" perturbations is fully characterized by a set of incentive adjustments or changes of probability measure that reweight types according to how much more or less resources a redistributive perturbation has to allocate to one type relative to their neighbor in order to preserve incentive compatibility. The optimal allocation then eliminates all redistributional arbitrage: any incentive-preserving, resource-feasible perturbation must lower expected welfare.

In the static environment considered by Mirrlees (1971) and Saez (2001), redistributional arbitrage leads to the same optimal tax implications as the so-called "ABC" formula (P. Diamond (1998)) which captures the efficiency vs. redistribution tradeoff. The static model equates consumption to after tax income, which implies that we can always use the income distribution to proxy for consumption, or vice versa. However, this stark implication of the static Mirrlees-Saez model is rejected by the data: consumption has a substantially thinner Pareto tail than income (Toda and Walsh (2015)), inconsistent with a static budget constraint equating consumption to after-tax income. The optimal top income tax drops from 80% in our preferred calibration to 51% if we instead use consumption to proxy for income inequality and replace the Pareto coefficient of the income distribution with the Pareto coefficient of the consumption distribution. In other words, Saez (2001)'s sufficient statistic representation of top optimal income taxes is based on an economic model that is inconsistent with the discrepancy between consumption and income inequality among top income earners.[2]

Allowing for multiple goods (or consumption and savings) and multiple margins of taxation breaks the tight link of consumption and after-tax income, and makes consumption and income data independently informative of the optimal tax design. Using them jointly leads to genuinely novel predictions, in line with the idea that we need two separate data sources for the identification of two separate tax distortions. We illustrate the predictive power of our approach by exploring the implications of redistributional arbitrage for income and savings taxes on top income earners.

We express the optimal top income and savings tax rates in terms of seven empirically observable statistics: the two Pareto tail coefficients of the distributions of taxable income and consumption; four elasticities that govern the consumption, savings, and labor supply responses to taxes, along with the preference complementarities between consumption and earnings; and the consumption share of income of high earners. We calibrate our model to match these statistics to their empirical counterpart and explore the quantitative implications of our formulas for top income and savings taxes.

Our formula for optimal income taxes coincides with that of Saez (2001) only if the utility function is linear in consumption, or if the savings share of income converges to zero for top income earners. If, as is the case in the data, neither of these conditions is satisfied, then the formula obtained in the static model overstates the top optimal labor income tax rate, both because it fails to account for the substitution between consumption and savings in response to tax changes, and because it fails to account for the fact that consumption is less unequally distributed than after-tax incomes in the data.

The formula of Saez (2001) instead determines the combined wedge on labor income and savings if the consumption share of income converges to zero for top income earners, a case that is implied by the thinner Pareto tail on consumption than on income. This is because with vanishing consumption shares at the top, the agents' incentive problem reduces again to a static tradeoff between leisure

[2]Consumption data provides an independent empirical test (and rejection) of the model underlying the representation of optimal taxes in the static model. This is an important caveat to the sufficient statistics approach: its implications rely on the empirical validity of the underlying economic model. The idea of using income and consumption data to test implications of efficient risk-sharing goes back to (at least) Townsend (1994). For applications of this idea to hidden information models see, e.g., Ligon (1998) and Kocherlakota and Pistaferri (2009).

and savings. However, that does not answer how the combined wedge should be broken up into an income and a savings wedge. While the savings wedge can, in principle, be positive or negative, simple calibrations suggest that it is optimal to shift some of the tax burden on top earners from income to savings, unless the consumption elasticity takes on an implausibly large value.

In quantitative terms, the formula of Saez (2001) suggested that top optimal income taxes may be very high, because high income inequality and low labor supply responses to income taxes for top earners generate high marginal benefits and low efficiency costs from tax distortions. By the same reasoning, our calibration suggest that the shift towards savings taxes can be very significant at the top of the income distribution, with savings taxes of up to 40%-50% of the level of savings and a corresponding reduction in top income taxes from a static optimum of 80% at our baseline calibration towards 60%. This shift from income towards savings taxes is a fairly robust feature of our quantitative results, and is driven by a combination of low consumption shares and/or thinner consumption tails at the top of the income distribution. These features of the data point suggest that the marginal benefit of redistributing consumption is small compared to the marginal benefit of redistributing savings. Hence it makes sense for the planner to shift part of the tax distortion towards savings at the top of the distribution.

While these optimal savings taxes may seem large, they do not strike us as implausible. In a life-cycle context with a 30-year gap between the working period and retirement and a 5% annual return on savings, a savings tax of 40% corresponds to a 1.8% annual tax on accumulated wealth, or a 35% capital income tax. These estimates are thus in the same ballpark as existing proposals of annual wealth taxes in the range of 1% to 2% (Saez and Zucman (2019a) and Saez and Zucman (2019b)). They also suggest that capital income should still be taxed at a significantly lower rate than labor income.[3]

Related Literature. Our paper relates to the optimal labor income and capital taxation literature originating with Mirrlees (1971) in the static setting, and Atkinson and Stiglitz (1976) in the dynamic setting, as well as the sufficient statistics approach towards estimating optimal tax rates that was pioneered by Saez (2001). By viewing tax policies as an arbitrage between different margins of redistribution, we generalize Saez' representation of optimal income taxes to a dynamic, or multiple-good, environment and derive a companion formula for optimal savings taxes. In linking this characterization of optimal taxes to its empirical counterparts, we show that optimal top

[3]Using a higher annualized return keeps the implied tax on wealth roughly constant, but translates into lower implied capital income taxes. Capital income taxes reach the range of 60%, i.e., roughly the same as labor income taxes, when annual returns are lowered to 3%.

taxes rely not only on labor income data – elasticities and Pareto coefficients – as in the canonical Saez (2001) framework – but also on consumption data. In particular, we rely on the analysis of Toda and Walsh (2015), who show that the Pareto tail of the distribution of consumption is significantly thinner than that of the income distribution. This observation turns out to have far-reaching consequences for how the burden of taxes should be shared between labor income and savings.

Our paper is related to a recent and growing literature that extends the sufficient statistics approach to optimal tax design in multi-good or multi-period settings (Golosov, Tsyvinski, and Werquin (2014), Scheuer and Werning (2016), Jacquet and Lehmann (2021), Spiritus, Lehmann, Renes, and Zoutman (2021), Ferey, Lockwood, and Taubinsky (2021), and Scheuer and Slemrod (2021)). In particular, Ferey, Lockwood, and Taubinsky (2021) and Scheuer and Slemrod (2021) are closest to our work. They arrive at similar quantitative conclusions out of a similar model, but using very different approaches.

Our paper differs from this literature in several important respects. First, these papers derive generalizations of the ABC formula and its analogue for savings by studying perturbations of the tax schedules. They thus remain focused on the tradeoff between incentives and redistribution. By contrast, our formulas for optimal taxes are based on redistributional arbitrage and bear little resemblance to the standard ABC expression. Conceptually, they highlight different – and, we believe, useful – insights than those found in the literature.

Second, this formula leads to a very simple analytical expression for the optimal income tax rates on high income earners, which generalizes the canonical static top tax rate result of Saez (2001) to a life-cycle environment, as well as an equally simple expression for top savings taxes, also expressed in terms of observable sufficient statistics. Because of our two margins of redistribution, the optimal tax rates depend naturally on the Pareto tail coefficients of both the consumption and the income distributions. Of the above papers, only Scheuer and Slemrod (2021) obtain a characterization of the capital tax rates on top earners, but taking as given the labor income tax. By contrast, we characterize the optimal labor and capital tax rates *jointly*. This joint optimization is important. For example, our characterization implies that the optimal savings tax does not depend on variables such as the Frisch elasticity. This comes as an immediate consequence of redistributional arbitrage between consumption and savings, but is obscured by alternative representations that explicitly link income and capital taxes.

Third, we show that the budget constraint in this class of models imposes over-identifying restrictions on the joint behavior of income, savings, and consumption of top earners. Calibrations of optimal tax formulas that do not satisfy these over-identifying restrictions are necessarily inconsistent with the underlying model within which they were derived. In particular, and in constrast to common practice, the corresponding Pareto coefficients cannot all be freely chosen from the data without verifying this consistency requirement. In our case, the Pareto coefficient on income must be equal to the minimum of those on consumption and savings. This motivates why our calibration focuses on consumption rather than wealth inequality as the natural empirical counterpart for evaluating the marginal benefits of redistributing consumption.

Saez and Stantcheva (2018) compute optimal top labor and capital tax rates under strong preference restrictions. In particular, ignoring income effects and the complementarity between consumption and earnings in preferences leads to labor income taxes identical to those in Saez (2001); by contrast, income effects and complementarities – and the use of consumption data to discipline them – play an important role in our model.[4]

Our dynamic model is based on Atkinson and Stiglitz (1976). Because we allow for arbitrary preferences, however, their uniform commodity taxation theorem only applies as special case of our framework. Christiansen (1984), Saez (2002), Jacobs and Boadway (2014), and Gauthier and Henriet (2018), among others, also generalize Atkinson and Stiglitz (1976) by allowing non-homothetic preferences. These papers typically constrain commodity or capital taxes to being linear. By contrast, our setting has an arbitrarily nonlinear tax system, which allows us to study the optimal tax rates on top earners. More broadly, a large literature extends Atkinson and Stiglitz (1976) in various directions from which we abstract: e.g., multidimensional heterogeneity (Mirrlees (1976), Saez (2002), Cremer, Pestieau, and Rochet (2003), P. Diamond and Spinnewijn (2011), Piketty and Saez (2013), Golosov, Troshkin, Tsyvinski, and Weinzierl (2013), Gahvari and Micheletto (2016), Saez and Stantcheva (2018), Gerritsen, Jacobs, Rusu, and Spiritus (2020), and Schulz (2021)) or uncertainty (P. A. Diamond and Mirrlees (1978), Golosov, Kocherlakota, and Tsyvinski (2003), Farhi and Werning (2010), Farhi and Werning (2013), and Golosov, Troshkin, and Tsyvinski (2016)). These papers often impose *a priori* restrictions on preferences – e.g., separable or GHH between consumption and labor with productivity types affecting only the disutility of effort. Besides being empirically relevant (Browning and Meghir (1991)), relaxing such restrictions turns out to be crucial in order to identify the parameters of our tax formula from the data: we show that such simple preferences are inconsistent with empirical evidence about the Pareto coefficients and elasticities necessary to calibrate optimal taxes.

Finally, our model solution, based on incentive-adjusted probability measures, builds on Hellwig

[4] Badel and Huggett (2017) derive optimal top income tax rate formulas in a model that allows for rich dynamics, general equilibrium effects, and fiscal externalities from a savings tax. However, all their applications take capital (or consumption) taxes as given and linear.

(2021) who developed these tools to study dynamic optimal taxation with non-separable preferences and generalized dynamic taxation results (inverse Euler equation, labor tax smoothing). We apply the same techniques to a simpler framework. This allows us to derive new insights on optimal income and savings taxes, and to confront these results to the data.

Outline of the Paper. We introduce our model in section 2 and derive our main results on optimal taxes as redistributional arbitrage in section 3. Section 4 discusses the comparison with the static tax formula of Saez (2001) and presents quantitative implications. Section 5 discusses possible extensions and generalizations of our model as well as additional implications of tax design by redistributional arbitrage. We conclude in Section 6.

2 The Model

2.1 Environment

There are two periods 0 and 1, and there is a measure-1 continuum of agents. At the beginning of period 0, each agent draws a stochastic type $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ i.i.d. according to c.d.f. $F(\cdot)$ and p.d.f. $f(\cdot)$. We refer to $H(\cdot) \equiv (1 - F(\cdot))/f(\cdot)$ as the inverse hazard rate associated with F. Agents' preferences are defined over consumption C and earnings Y in period 0, and consumption S ("savings") in period 1. They are represented as

$$U(C, Y; \theta) + \beta V(S)$$

where U is twice continuously differentiable, $U_C > 0$, $U_{CC} < 0$, $U_Y < 0$, $U_{YY} < 0$, $U_\theta > 0$ and U otherwise satisfies the usual Inada conditions as C or Y approach 0 or ∞. The function V is twice continuously differentiable, increasing and concave in S and satisfies the usual Inada conditions as S approaches 0 or ∞. Consumption, earnings, and savings are assumed to be observable but individual types are the agents' private information. Resources can be saved at a rate R from period 0 to 1.

We make two additional assumptions about $U(C, Y; \theta)$. The first is the *Strict Single-Crossing Condition:* $-U_Y(C, Y; \theta)/U_C(C, Y; \theta)$ is strictly decreasing in θ for all $(C, Y; \theta)$, or

$$\frac{U_{C\theta}}{U_C} - \frac{U_{Y\theta}}{U_Y} > 0.$$

This assumption guarantees monotonicity of any incentive-compatible allocation: on the margin, higher types are more willing to work.

The second assumption concerns the signs of $U_{C\theta}$ and $U_{Y\theta}$, which determine redistribution motives in the optimal tax problem. In the canonical Mirrlees model, $U(C,Y;\theta) = \mathcal{U}(C, Y/P(\theta))$ where $P(\theta)$ represents the productivity or ability of type θ, and $Y/P(\theta)$ represents hours worked; thus, $U_{Y\theta}/U_Y < 0$.[5] More broadly we assume that $U_{Y\theta}/U_Y < 0$ and interpret this assumption as giving rise to a redistribution motive of effort from less to more productive agents, or equivalently, of leisure towards less productive agents – that is, redistribution *"from each according to their ability"*.

But the type θ may also directly enter the marginal utility of consumption when $U_{C\theta} \neq 0$. This results in a second redistribution motive of consumption towards those agents who have the highest marginal utilities or "consumption needs" – that is, redistribution *"to each according to their needs"*. We do not impose any specific restrictions on $U_{C\theta}/U_C$, but we assume that it does not change sign and is either non-positive or non-negative everywhere. In the former case, both redistribution motives favor lower types. In the latter, consumption needs are higher for higher types, in which case the two redistribution motives are not aligned. Nevertheless, the single-crossing condition guarantees that it is always optimal to redistribute from higher to lower types, i.e. the planner has a motive of demanding higher effort from, and offering higher consumption to high types.

Elasticity Concepts. We introduce several important elasticities. Let

$$\mathcal{E}_Y(\theta) \equiv \frac{\partial \ln(-U_Y/U_C)}{\partial \ln Y} = \frac{YU_{YY}}{U_Y} - \frac{YU_{CY}}{U_C}$$

and

$$\mathcal{E}_C(\theta) \equiv \frac{\partial \ln(-U_Y/U_C)}{\partial \ln C} = \frac{-CU_{CC}}{U_C} - \frac{CU_{CY}}{-U_Y}$$

be the elasticities of the agent's marginal rate of substitution with respect to earnings and consumption, respectively. Moreover, let

$$\mathcal{E}_{CY}(\theta) \equiv \frac{\partial \ln U_C}{\partial \ln Y} = \frac{YU_{CY}}{U_C}$$

denote the complementarity between consumption and earnings (or labor), and

$$\mathcal{E}_S(\theta) \equiv -\frac{\partial \ln V'(S)}{\partial \ln S} = -\frac{SV''(S)}{V'(S)}$$

[5] While it is convenient for the analysis to define preferences in terms of the observables C, Y, and S, it is straightforward to map the type-contingent preference over earnings into a preference over leisure or hours worked.

denote the coefficient of relative risk aversion in the second period. These four variables $\mathcal{E}_C$, $\mathcal{E}_Y$, $\mathcal{E}_{CY}$, $\mathcal{E}_S$ play a key role in our analysis as they jointly determine the strength of income and substitution effects on labor supply and consumption in both periods. We assume that all four elasticities are non-negative over their entire support, which implies that leisure, consumption and savings are normal goods, and consumption and leisure are net substitutes. In Section 4 we relate these elasticities to empirically observable parameters.

2.2 Social Planner's Problem

The utilitarian social planner's optimal allocation $\{C(\theta), Y(\theta), S(\theta)\}$ minimizes the net present value of transfers

$$\int_{\underline{\theta}}^{\overline{\theta}} \left(C(\theta) - Y(\theta) + R^{-1} S(\theta)\right) f(\theta)\, d\theta$$

subject to the promise-keeping constraint

$$\int_{\underline{\theta}}^{\overline{\theta}} \{U(C(\theta), Y(\theta); \theta) + \beta V(S(\theta))\} f(\theta)\, d\theta \;\geq\; v_0,$$

and the incentive compatibility constraint

$$U(C(\theta), Y(\theta); \theta) + \beta V(S(\theta)) \;\geq\; U\left(C(\theta'), Y(\theta'); \theta\right) + \beta V\left(S(\theta')\right)$$

for all types θ and annoucements θ'. If the utility promise v_0 is chosen so that the net present value of transfers at the optimum equals 0, the solution to the problem corresponds to the allocation that maximizes the expected utility of agents, subject to satisfying an aggregate break-even condition. But this general dual formulation allows us to vary the marginal value of public funds by varying the initial expected utility promise v_0.

We solve this problem using a Myersonian approach, replacing full incentive compatibility by local incentive compatibility. Define the indirect utility function $W(\theta) \equiv U(C(\theta), Y(\theta); \theta) + \beta V(S(\theta))$. Then an allocation is locally incentive-compatible, if it satisfies

$$W'(\theta) = U_\theta(C(\theta), Y(\theta); \theta).$$

We refer to $U_\theta(\theta) \equiv U_\theta(C(\theta), Y(\theta); \theta)$ as the *marginal information rent* of type θ. The solution to this relaxed problem is obtained using optimal control techniques and is fully described in the Appendix. To ease notation, we further write $X(\theta) \equiv X(C(\theta), Y(\theta), S(\theta); \theta)$ for any function X

of both the allocation $(C(\theta), Y(\theta), S(\theta))$ and the type θ at date 1.

Let $\tau_Y(\theta) \equiv 1 + U_Y(\theta)/U_C(\theta)$ denote the labor wedge at θ, i.e., the intra-temporal distortion between the marginal product and the marginal rate of substitution between consumption and earnings. Let $\tau_S(\theta) \equiv \beta R V'(\theta)/U_C(\theta) - 1$ denote the savings wedge at θ, i.e., the inter-temporal distortion in the agent's first-order condition for savings. Below, we first characterize the optimal wedges $\{\tau_Y(\theta), \tau_S(\theta)\}$ implied by the optimal allocation $\{C(\cdot), Y(\cdot), S(\cdot)\}$ that solves the planner's problem, and then discuss how to decentralize these allocations through the appropriate design of income and savings taxes. In particular, we can show that there exists a decentralization in which $\{\tau_Y(\cdot), \tau_S(\cdot)\}$ map one-for-one into marginal income and savings taxes at an optimal tax system. Note finally that while $\tau_Y(\cdot)$ represents a marginal labor income tax on gross earnings, $\tau_S(\cdot)$ represents the savings wedge as a proportion of net savings $S(\cdot)$. For constant top savings wedges, this translates into a top marginal tax on gross savings that is equal to $\tau_S(\cdot)/(1+\tau_S(\cdot))$.

2.3 Incentive-Adjusted Probabilities

Our representation of the optimality conditions borrows from Hellwig (2021) who translates non-separability in preferences into a change in probability measures, i.e., an incentive-adjustment to the probabilities, or planner weights, that are used to evaluate marginal changes to consumption and earnings allocations. For a given allocation $\{C(\cdot), Y(\cdot), S(\cdot)\}$, we define

$$\widehat{M}(\theta) = \frac{\widehat{m}(\theta)}{U_C(\theta)} \quad \text{where} \quad \widehat{m}(\theta) = e^{-\int_{\theta}^{\overline{\theta}} \frac{U_{\theta C}(\theta')}{U_C(\theta')} d\theta'}$$

and

$$\widetilde{M}(\theta) = \frac{\widetilde{m}(\theta)}{-U_Y(\theta)} \quad \text{where} \quad \widetilde{m}(\theta) = e^{-\int_{\theta}^{\overline{\theta}} \frac{U_{\theta Y}(\theta')}{U_Y(\theta')} d\theta'}.$$

Define the incentive-adjusted probability measures

$$\widehat{f}(\theta) \equiv \frac{f(\theta)\,\widehat{m}(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} f(\theta')\,\widehat{m}(\theta')\,d\theta'} \quad \text{and} \quad \widetilde{f}(\theta) \equiv \frac{f(\theta)\,\widetilde{m}(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} f(\theta')\,\widetilde{m}(\theta')\,d\theta'}$$

and let $\widehat{F}(\cdot)$ and $\widetilde{F}(\cdot)$ denote the cumulative distribution functions, $\widehat{H}(\cdot)$ and $\widetilde{H}(\cdot)$ the corresponding inverse hazard rates, and $\widehat{\mathbb{E}}(\cdot)$ and $\widetilde{\mathbb{E}}(\cdot)$ the expectation operators, associated with these incentive-adjusted distributions $\widehat{f}(\theta)$ and $\widetilde{f}(\theta)$.

These incentive-adjustments represent how the planner must reweight the marginal allocation of consumption and leisure (or earnings) so as to preserve local incentive compatibility constraints. Both incentive-adjustments decompose into a risk component $U_C(\theta)$ or $-U_Y(\theta)$ that captures the redistribution motive, and an incentive component $\widehat{M}(\theta)$ or $\widetilde{M}(\theta)$ that represents the adjustment to the marginal redistribution of resources. As we discuss in the next subsection, a local perturbation to an allocation preserves incentive compatibility if and only if consumption is redistributed in proportion to $\widehat{M}(\cdot)$, and earnings or leisure in proportion to $\widetilde{M}(\cdot)$.

The direction of these incentive-adjustments depends on the signs of $U_{\theta C}$ and $U_{\theta Y}$ in a way that is naturally linked to the interpretation of $U_{\theta C}/U_C$ and $U_{\theta Y}/U_Y$ as needs- and ability-based redistribution motives. This allows us to rank the distributions by first-order stochastic dominance. Since $U_{\theta Y}/U_Y < 0$, increasing utilities by reducing earnings (i.e., increasing leisure) lowers information rents, and thus allows for more redistribution of utility towards lower types. Therefore $\widetilde{F}(\cdot)$ overweighs the lower types relative to $F(\cdot)$.

Likewise, $\widehat{F}(\cdot)$ is shifted towards higher (lower) types, whenever higher types have higher (lower) consumption needs. When higher types have lower consumption needs ($U_{\theta C} < 0$), increasing consumption reduces information rents and thus allows for more redistribution towards lower types. By contrast, when consumption needs are increasing in type, extra consumption increases information rents and reduces redistribution towards lower types, in which case $\widehat{F}(\cdot)$ first-order stochastically dominates $F(\cdot)$. Finally, the single-crossing condition insures that $\widehat{F}(\cdot)$ first-order stochastically dominates $\widetilde{F}(\cdot)$, i.e., that the effect of earnings changes on marginal information rents outweights the effect of consumption changes, thus resulting in more progressive redistribution of leisure than of consumption, on the margin.

2.4 Optimal Allocations

Suppose that the optimal allocation is strictly positive everywhere. Using the above characterization of incentive-adjusted probabilities, we provide the following representation of optimal allocations:

Proposition 1. *The optimal allocation satisfies the following set of optimality conditions*

$$\frac{\tau_Y(\theta)}{1-\tau_Y(\theta)}\frac{1}{U_C(\theta)} = A(\theta)\cdot\frac{B_C(\theta)}{U_C(\theta)} = A(\theta)\cdot\frac{B_Y(\theta)}{-U_Y(\theta)} = A(\theta)\cdot\frac{B_S(\theta)}{V'(S(\theta))}$$

where

$$A(\theta) = \frac{U_{\theta C}(\theta)}{U_C(\theta)} - \frac{U_{\theta Y}(\theta)}{U_Y(\theta)}$$

and

$$
\begin{aligned}
B_C(\theta) &= \widehat{H}(\theta)\left(\widehat{\mathbb{E}}\left[\frac{U_C(\theta)}{U_C(\theta')}|\theta' \geq \theta\right] - \lambda U_C(\theta)\right) \\
B_Y(\theta) &= \widetilde{H}(\theta)\left(\widetilde{\mathbb{E}}\left[\frac{-U_Y(\theta)}{-U_Y(\theta')}|\theta' \geq \theta\right] - \lambda(-U_Y(\theta))\right) \\
B_S(\theta) &= H(\theta)\left(\mathbb{E}\left[\frac{V'(S(\theta))}{V'(S(\theta'))}|\theta' \geq \theta\right] - \lambda\beta R V'(S(\theta))\right)
\end{aligned}
\tag{1}
$$

with

$$
\lambda = \widehat{\mathbb{E}}\left[\frac{1}{U_C(\theta')}\right] = \widetilde{\mathbb{E}}\left[\frac{1}{-U_Y(\theta')}\right] = \frac{1}{\beta R}\mathbb{E}\left[\frac{1}{V'(S(\theta'))}\right].
$$

Moreover, if savings are unbounded above and $\lim_{\theta\to\bar{\theta}} \tau_Y(\theta) < 1$, *then optimal allocations satisfy the Inada condition* $\lim_{\theta\to\bar{\theta}} U_C(\theta) = \lim_{\theta\to\bar{\theta}}(-U_Y(\theta)) = \lim_{\theta\to\bar{\theta}} V'(S(\theta)) = 0$.

The first condition re-states and generalizes the well-known ABC-formula from Proposition 1 in Saez (2001) to the present environment with consumption and savings. The second and third conditions provide a complement, based on the same logic as the ABC-formula, but focusing on the optimal redistribution of leisure and savings rather than consumption. These conditions are obtained from the first-order conditions to the planner's problem along the same lines as the ABC-formula but by focusing on allocation of leisure and savings rather than consumption.

Corollary 1. *The optimal labor wedge* $\tau_Y = 1 + U_Y/U_C$ *admits the following three (equivalent) representations:*

$$
\frac{\tau_Y(\theta)}{1-\tau_Y(\theta)} = A(\theta)\cdot B_C(\theta), \qquad \tau_Y(\theta) = A(\theta)\cdot B_Y(\theta)
$$

and

$$
1-\tau_Y(\theta) = \frac{B_Y(\theta)}{B_C(\theta)} \tag{2}
$$

In addition, the optimal savings wedge $1+\tau_S = \beta R V'/U_C$ *is uniquely characterized by:*

$$
1+\tau_S(\theta) = \frac{B_S(\theta)}{B_C(\theta)}. \tag{3}
$$

Proposition 1 and Corollary 1 summarize the principle of redistributional arbitrage: at the optimal allocation, the marginal cost of efficiency distortions, which is given by $\frac{\tau_Y}{1-\tau_Y}A^{-1}/U_C$, must be equal to the marginal benefits of redistributing consumption, leisure or savings, which are given respectively by B_C/U_C, $B_Y/(-U_Y)$, and B_S/V'. The planner thus equates marginal costs and benefits of redistribution, and is indifferent between redistribution via consumption or via earnings at each realization of θ.

Corollary 1 shows that the optimal labor wedge can be represented in three different manners. The first is a restatement of the ABC formula. The second is the analogue based on equating the marginal cost of efficiency distortions to the marginal benefit of redistributing leisure. The third condition is obtained from the first two by eliminating A; it states that the marginal benefits of redistributing consumption and leisure must be equalized. Importantly, since leisure, consumption and savings are separately linked to each other through the incentive compatibility and budget constraints, these three conditions are all equivalent to each other. However, as we shall see below, the three expressions differ in terms of the observable statistics that they emphasize, and therefore the calibration of optimal income taxes. As a consequence of this triple representation of the optimal labor wedge, the model imposes testable restrictions on observables (income, consumption and savings) that must be empirically confirmed before using the model to estimate optimal income taxes.

By contrast, the optimal savings wedge is uniquely pinned down by the equalization of the marginal benefits of redistributing consumption and savings. This observation is interesting for two reasons. First, it shows that optimal income and savings taxes can be represented through the same lens as the result of redistributional arbitrage. Second, this representation of the optimal savings tax, in contrast to the existing literature, is independent of the optimal income tax. This substantially simplifies the characterization of the optimal savings tax, and has direct implications for the set of parameters and observables that determine the optimal savings wedge: the optimal savings wedge should only depend on parameters that enter B_S and B_C directly, but is independent of parameters that only affect B_Y or A without affecting B_S and B_C. In particular, if preferences are additively separable between consumption, leisure and savings, then equation (3) implies that the optimal savings tax should be independent of parameter choices regarding the discount factor β, the rate of return R, or parameter choices regarding leisure such as the Frisch elasticity of labor supply.

Contrast this with the alternative ABC representation of the optimal savings tax:

$$1 + \tau_S(\theta) = \beta R \cdot A(\theta) \cdot \frac{1 - \tau_Y(\theta)}{\tau_Y(\theta)} B_S(\theta).$$

This representation clearly suggests a dependence of the optimal savings tax on β, R, and the parameters that determine A and τ_Y. This dependence must therefore be spurious: the structure of the model imposes a tight overidentifying relationship between these parameters. As a consequence, it would be incorrect to calibrate the ABC formula for, say, the capital tax rate by choosing these parameters freely and independently from the data. For instance, running robustness checks of such a calibration by varying the Frisch elasticity independently of the other parameters of the formula would be inconsistent with the model within which this optimal tax formula was derived.

Perturbation-Based Interpretation of Proposition 1. The interpretation of these expressions as marginal costs and benefits of redistribution stems from a simple set of perturbation arguments. Consider a perturbation $(\Delta C_1(\theta), \Delta Y_1(\theta))$ that leaves a given type θ's utility unchanged, while marginally increasing this type's output. This perturbation raises resources in proportion to $\frac{\tau_Y}{1-\tau_Y}\frac{1}{U_C}$ in terms of consumption goods. At the same time, the perturbation increases the marginal information rent U_θ at θ by $\frac{U_{\theta C}}{U_C} - \frac{U_{\theta Y}}{U_Y} > 0$. Therefore, upon reducing the distortion at θ, the planner cannot freely redistribute the extra resources across all types, but must raise the utility of all types $\theta' > \theta$ by an extra ΔU_θ, relative to all types $\theta' < \theta$. Hence, at each θ, the planner faces a simple trade-off between efficiency and redistribution: More redistribution around θ must come at the cost of lower efficiency at θ, and vice versa. The term A then represents the marginal change in information rents $\frac{U_{\theta C}}{U_C} - \frac{U_{\theta Y}}{U_Y}$, and the expression $\frac{\tau_Y}{1-\tau_Y}A^{-1}/U_C$ the marginal rate of substitution between incentives and redistribution, or the marginal efficiency cost of additional redistribution around θ.

The terms B_C/U_C, $B_Y/(-U_Y)$, and B_S/V' represent how much the planner values additional redistribution locally around a given θ. Here, B_C/U_C measures the marginal value of redistribing consumption, $B_Y/(-U_Y)$ the marginal value of redistributing leisure, and B_S/V' the marginal value of redistributing savings. These interpretations are based on three elementary perturbations that provide further intuition for the incentive adjustment to probability measures.

Suppose that the planner wishes to redistribute consumption from types $\theta' > \theta$ to types $\theta' < \theta$, while keeping expected utility unchanged and preserving incentive compatibility for all $\theta' \neq \theta$. Let ΔC, with $\Delta C < 0$ for $\theta' > \theta$ and $\Delta C > 0$ for $\theta' < \theta$ denote a small perturbation to consumption for all $\theta' \neq \theta$. This perturbation changes utility by $\Delta W = U_C \Delta C$ and marginal information rents by $\Delta U_\theta = U_{\theta C}\Delta C$. It therefore preserves local incentive compatibility if and only if

$$\Delta W' = \Delta U_\theta = \frac{U_{\theta C}}{U_C}\Delta W.$$

The unique perturbation that preserves local incentive compatibility for all $\theta' \neq \theta$ then redistributes consumption in proportion to $\widehat{M}$ and utility in proportion to $\widehat{m}$: For θ' just above θ, the perturbation changes U_θ by $\Delta U_\theta = (U_{\theta C}/U_C) \cdot \Delta W$, and this adjustment to the slope must be offered to all $\theta'' > \theta'$ to restore local incentive compatibility at θ'. But these modifications for any $\theta'' > \theta'$ then generate further adjustments for all $\theta''' > \theta''$, and so on. By re-weighting the utility perturbations according to $\widehat{m}$, the perturbation preserves local incentive compatibility for all types: the ODE is solved by "integrating up" the cumulative utility changes for higher types that are required as a result of preserving local incentive compatibility at all lower types. By construction, $\widehat{m}$ is set so that $\frac{\widehat{m}'}{\widehat{m}} = \frac{U_{C\theta}}{U_C}$, i.e., the rate of change embedded in the incentive adjustment is set equal to the impact of the consumption perturbation on type θ's marginal information rent.

We can then set the change in utilities around θ so that expected utility remains constant, and calculate the expected gain in resources $\mathbb{E}[\Delta C]$ associated with this perturbation. This resource gain is always positive if higher types have lower marginal utilities of consumption. Simple algebra shows that $\mathbb{E}[\Delta C] = \delta f(\theta) \frac{B_C}{U_C}$, where δ represents the change in utility at θ that is implemented by the perturbation, and $f(\theta)$ the density of types at θ, which scales the marginal cost of efficiency distortions. Therefore B_C represents the marginal value of redistribution around θ via consumption, i.e., the marginal resource gain from transfering consumption from types $\theta' > \theta$ to types $\theta' < \theta$ while maintaining incentive compatibility and keeping expected utility unchanged, weighted by the density of types at θ.

Along the same lines, a perturbation that redistributes leisure or earnings around θ while preserving local incentive compatibility and expected-utility must reweight states according to $\widetilde{m}$, resulting in the above expression B_Y as the marginal value of redistributing leisure. The leisure-based incentive adjustment $\widetilde{m}$ is constructed so that $\frac{\widetilde{m}'}{\widetilde{m}} = \frac{U_{Y\theta}}{U_Y}$, again equating its rate of change to the impact of perturbing leisure on type θ's marginal information rent.

Finally, the perturbation that redistributes savings around θ while preserving local incentive compatibility and expected-utility is simplified by the fact that the marginal utility of savings is independent of θ. Hence the perturbation that preserves incentive-compatibility and expected utility redistributes savings in proportion to $1/V'$, and no incentive adjustment to probabilities is necessary.

These four elementary perturbations can also be used to identify possible welfare improvements. If the marginal value of redistribution is higher for consumption than for leisure, the planner gains resources by redistributing more consumption goods, but reducing redistribution of leisure. On the other hand, if one of the marginal benefits of redistribution exceeds the marginal cost of efficiency distortions, increasing taxes and redistribution locally is welfare improving. Hence these elementary perturbations also offer basic guidance on the directions of improvement to a sub-optimal tax schedule.

2.5 Decomposing the Marginal Benefits of Redistribution

From now on we focus exclusively on the representations of optimal labor and savings wedges that are based on redistributional arbitrage, i.e., equations (2) and (3). The three expressions (1) show that each of these marginal benefits $B_Y(\theta), B_C(\theta), B_S(\theta)$ decomposes into two terms.

The terms $\lambda(-U_Y)$, λU_C and $\lambda\beta RV'$ capture the welfare losses of reducing income, consumption or savings of type θ and redistributing the resulting resources across the population. They represent an inverse Pareto weight on type θ, where λ represents the shadow cost of increasing expected utility by increasing leisure, consumption or savings, subject to preserving incentive compatibility; at the optimal allocation, these three expected shadow costs have to be equal. The Inada condition of Proposition 1 implies that these terms all converge to zero when allocations are unbounded at the top and marginal utilities converge to zero.

Top income and savings taxes are thus determined by the ratios of

$$\widetilde{H}(\theta)\,\widetilde{\mathbb{E}}\left[\frac{-U_Y(\theta)}{-U_Y(\theta')}|\theta' \geq \theta\right], \quad \widehat{H}(\theta)\,\widehat{\mathbb{E}}\left[\frac{U_C(\theta)}{U_C(\theta')}|\theta' \geq \theta\right], \quad H(\theta)\,\mathbb{E}\left[\frac{V'(S(\theta))}{V'(S(\theta'))}|\theta' \geq \theta\right].$$

These ratios describe the trade-off between redistributing resources from the top via earnings, via consumption or via savings – or in other words, how the social planner maximizes the extraction of resources from the top earners by asking them to work more, consume less, or save less. This is precisely the trade-off underlying redistributional arbitrage at the top of the income distribution.

To interpret these expressions, consider a perturbation by which the planner marginally increases consumption for type θ by δ_C. The infra-marginal changes to consumption for higher types must be proportional to $\widehat{M}(\theta')/\widehat{M}(\theta)$ to preserve incentive compatibility. Integrating up, the total resource cost of increasing consumption of type θ by δ_C is equal to $\mathbb{E}[\widehat{M}(\theta')/\widehat{M}(\theta)|\theta' \geq \theta]\cdot\delta_C$. Likewise, the total resource gain of raising earnings of type θ by δ_Y is $\mathbb{E}[\widetilde{M}(\theta')/\widetilde{M}(\theta)|\theta' \geq \theta]\cdot\delta_Y$. Now redistributional arbitrage should dictate that if δ_Y and δ_C are set so that type θ is kept exactly indifferent, i.e. $\delta_C = U_C(\theta)\,\delta$ and $\delta_Y = -U_Y(\theta)\,\delta$, then the resource gain from raising more revenue at the top of the income distribution should be exactly equal to the cost of providing the top types with higher consumption. Applying the definitions of the incentive adjusted probability measures, the two expressions can be rewritten as $H(\theta)\,\mathbb{E}[\widehat{M}(\theta')/\widehat{M}(\theta)|\theta' \geq \theta] = \widehat{H}(\theta)\,\widehat{\mathbb{E}}[U_C(\theta)/U_C(\theta')|\theta' \geq \theta]$ and $H(\theta)\,\mathbb{E}[\widetilde{M}(\theta')/\widetilde{M}(\theta)|\theta' \geq \theta] = \widetilde{H}(\theta)\,\widetilde{\mathbb{E}}[U_Y(\theta)/U_Y(\theta')|\theta' \geq \theta]$.

This leads to a reinterpretation of optimal wedges at the top of the income distribution: the optimal tax rate equates the agent's marginal rate of substitution, i.e., the agent's private trade-off between earnings and consumption, to the social infra-marginal tradeoff in redistribution from the top via earnings and consumption.

We apply the same arguments to the expression for B_S and the expression for the optimal savings wedge. Thus, we can state the optimal savings distortion as resulting from a trade-off between benefits of redistributing current consumption and redistributing savings around a given type θ: the agent's inter-temporal marginal rate of substitution between consumption and savings must be equal to the planner's infra-marginal trade-off between consumption and savings for all $\theta' > \theta$.

2.6 When Should Savings Be Taxed?

The uniform commodition taxation theorem of Atkinson and Stiglitz (1976) is nested as a special case of our savings wedge representation (3). In the context of our model, the Atkinson-Stiglitz theorem says that the optimal savings wedge must be equal to zero for all types, if and only if the marginal utility of consumption is independent of types θ; that is, if preferences are separable between consumption and income, and the utility of consumption is homogeneous across consumers. If this is the case, the optimal allocation always equalizes the marginal benefit of redistributing via savings to the marginal benefit of redistributing via consumption, and there is no reason to tax savings in addition to income. The following corollary also shows that the converse statement is true.

Corollary 2. *The optimal allocation satisfies $B_S(\theta) \gtreqless B_C(\theta)$ and the optimal savings wedge is $\tau_S(\theta) \gtreqless 0$ for all θ, if and only if $U_{C\theta}(\theta) \lesseqgtr 0$ for all θ.*

In other words, the optimal savings tax inherits the sign of $-U_{\theta C}$. If the marginal utility is increasing with θ, so that more productive types also have higher consumption needs, then it is optimal to subsidize savings at the top of the income distribution. If instead consumption needs (i.e., the marginal utility) is decreasing with θ, then it is optimal to tax savings at the top of the income distribution. Hellwig (2021) establishes the same result in a general multi-period dynamic Mirrlees model and shows that the alignment of ability with consumption needs offers a new rationale for taxing or subsidizing savings.

Coupled with the fact that preferences over savings are independent across types, the condition $U_{C\theta} = 0$ for all θ implies that the agents' marginal rate of substitution between consumption and savings is independent of θ. This is equivalent to the weak separability assumption imposed in Atkinson and Stiglitz (1976). Notice that $B_S = B_C$ is equivalent to

$$\mathbb{E}\left[\frac{1}{\beta R V'(S(\theta'))} | \theta' \geq \theta\right] = \widehat{\mathbb{E}}\left[\frac{1}{U_C(\theta')} | \theta' \geq \theta\right], \quad \forall \theta.$$

When preferences are such that $U_{C\theta} = 0$ for all θ, then $\widehat{\mathbb{E}} = \mathbb{E}$, i.e., the incentive adjustment for consumption disappears; it is then straight-forward to check that the latter condition is equivalent to $1/U_C = 1/\beta RV'$, or $\tau_S = 0$ for all θ.

Why does the uniform taxation principle depend on this incentive adjustment? The intuition for this result is as follows: A perturbation that reduces consumption for type θ by δ in order to increase savings lowers the current utility by $U_C\delta$ and changes the type's information rent by $U_{\theta C}\delta$. The ratio $U_{\theta C}/U_C$ thus measures the ratio of the change in information rents to the change in utility that comes with an increase in savings. If additional savings reduce information rent ($U_{\theta C} > 0$), then this allows the planner to increase the static redistribution from higher towards lower types, which leads to a rationale for subsidizing savings. If instead the marginal savings increase the agent's information rent ($U_{\theta C} < 0$), then savings reduce the scope for static redistribution, which makes it optimal to tax savings.[6]

3 Optimal Top Tax Rates: A Sufficient Statistics Representation

3.1 Sufficient Statistics for B_C, B_Y, and B_S

In this section, we show how to map the incentive-adjustments $\widehat{M}$ and $\widetilde{M}$, and hence our change of measure, to observable statistics of the distribution of earnings and consumption and some key elasticity parameters. This in turn allows us to express B_C, B_Y, and B_S, and hence the top income and savings taxes, in terms of observable statistics and estimated elasticities.

Proposition 2. *$\widehat{M}(\cdot)$ and $\widetilde{M}(\cdot)$ take the form*

$$\begin{aligned}\widehat{M}(\theta) &= e^{-\int_{\theta}^{\overline{\theta}} \mathcal{E}_C(\theta')\frac{C'(\theta')}{C(\theta')}d\theta'}\Psi(\theta)\\ \widetilde{M}(\theta) &= e^{\int_{\theta}^{\overline{\theta}} \mathcal{E}_Y(\theta')\frac{Y'(\theta')}{Y(\theta')}d\theta'}\Psi(\theta)\end{aligned}$$

with

$$\Psi(\theta) = e^{\int_{\theta}^{\overline{\theta}}(1-\kappa(\theta'))\mathcal{E}_{CY}(\theta')\frac{Y'(\theta')}{Y(\theta')}d\theta'},$$

and

$$V'(S(\theta)) = e^{\int_{\theta}^{\overline{\theta}} \mathcal{E}_S(\theta')\frac{S'(\theta')}{S(\theta')}d\theta'}$$

[6] The intuition and the result generalizes to preferences of the form $U(C, S, Y, \theta)$, allowing for interaction between S and θ along the same lines as C and θ. Uniform commodity taxation then holds ($\tau_S = 0$ for all θ) if and only if $\frac{U_{\theta C}}{U_C} = \frac{U_{\theta S}}{U_S}$ for all θ, in which case the incentive-adjusted probability measures for C and S are the same.

where $\mathcal{E}_C(\cdot), \mathcal{E}_Y(\cdot), \mathcal{E}_{CY}(\cdot), \mathcal{E}_S(\cdot)$ are defined as above, and where $\kappa(\theta) \equiv \frac{C'(\theta)}{(1-\tau_Y(\theta))Y'(\theta)}$ represents the marginal increase in consumption, relative to the marginal increase in after-tax income, induced by a marginal increase in θ.[7]

Suppose first that the utility function is separable, $U_{CY} = 0$, in which case $\mathcal{E}_{CY}(\cdot) = 0$ and $\Psi(\cdot) = 1$. Then, the incentive component $\widehat{M}(\cdot)$ only depends on the distribution of consumption and the elasticity $\mathcal{E}_C(\cdot)$ of the marginal rate of substitution with respect to consumption. $\widehat{M}(\cdot)$ is non-decreasing in θ, so that on the margin the redistribution of consumption must be regressive. Likewise, $\widetilde{M}(\cdot)$ only depends on the distribution of earnings and the elasticity $\mathcal{E}_Y(\cdot)$ of the marginal rate of substitution with respect to earnings. $\widetilde{M}(\cdot)$ is decreasing in θ, i.e., the redistribution of leisure is progressive. Furthermore, if $\mathcal{E}_C(\cdot)$ and $\mathcal{E}_Y(\cdot)$ converge to constants equal to $\mathcal{E}_C$ and $\mathcal{E}_Y$ for high θ, then $\widehat{M}(\theta) \sim C(\theta)^{\mathcal{E}_C}$ and $\widetilde{M}(\theta) \sim Y(\theta)^{-\mathcal{E}_Y}$. Hence, the upper tail of the distribution of earnings (resp., consumption) and the elasticity $\mathcal{E}_Y$ (resp., $\mathcal{E}_C$) are sufficient to estimate the first term in B_Y (resp., B_C).

If $U_{CY} \neq 0$ then $\widehat{M}(\theta) \sim C(\theta)^{\mathcal{E}_C} \Psi(\theta)$ and $\widetilde{M}(\theta) \sim Y(\theta)^{-\mathcal{E}_Y} \Psi(\theta)$, i.e., the term $\Psi(\theta)$ corrects these incentive adjustments for preference complementarity between consumption and earnings. If $\mathcal{E}_{CY}(\cdot)$ and $\kappa(\theta)$ converge to finite limits $\mathcal{E}_{CY}$ and $\kappa \in (0,1)$ at the top of the type distribution, then $\Psi(\theta) \sim Y(\theta)^{-(1-\kappa)\mathcal{E}_{CY}}$. Below we show that we can use the MPC of top earners, or their consumption share of income, to derive an empirical counterpart for κ, and hence $\Psi(\cdot)$.

Finally, the first term in B_S can analogously be mapped to the upper tail of the distribution of savings, along with the savings elasticity $\mathcal{E}_S(\cdot)$, thus leading to the last equation of the proposition. If $\mathcal{E}_S(\cdot)$ converges to a constant, we then obtain $V'(S(\theta)) \sim S(\theta)^{-\mathcal{E}_S}$.

Proposition 2 shows that with two consumption goods, two different distributions are required to fully infer the shape of the incentive adjustment, and more generally the properties of optimal income taxes. In the sequel, we focus on consumption and income as the two observables, with savings being determined as the residual from the agents' inter-temporal budget constraint.

3.2 Optimal Top Tax Rates

It is now straight-forward to compute optimal labor and savings wedges at the top of the income distribution, using the above characterizations in terms of the distributions of earnings, consumption

[7] $\Psi(\theta)$ can equivalently be represented more symmetrically as $\exp \int_{\theta}^{\overline{\theta}} \left[\mathcal{E}_{CY}(\theta') \frac{Y'(\theta')}{Y(\theta')} - \mathcal{E}'_{CY}(\theta') \frac{C'(\theta')}{C(\theta')} \right] d\theta'$ where $\mathcal{E}'_{CY}$ is another coefficient of complementarity, defined by $\frac{CU_{CY}}{-U_Y}$. The variable κ allows us to express Ψ in terms of the standard elasticity variable $\mathcal{E}_{CY} \equiv \frac{YU_{CY}}{U_C}$ only by using the first order condition $1 - \tau_Y = \frac{-U_Y}{U_C}$. This is a useful transformation since empirical estimates of κ and $\mathcal{E}_{CY}$ are more readily available in the literature than $\mathcal{E}'_{CY}$, as we argue in Section 4.

and savings. In particular, consider the following assumption:

Assumption 1. *The optimal allocation* $\{C(\cdot), Y(\cdot), S(\cdot)\}$ *is co-monotonic, and the distributions of earnings, consumption and savings have unbounded support and upper Pareto tails with coefficients* ξ_Y*,* ξ_C *and* ξ_S*. In addition, the elasticities* $\mathcal{E}_C(\cdot)$*,* $\mathcal{E}_Y(\cdot), \mathcal{E}_{CY}(\cdot)$*,* $\mathcal{E}_S(\cdot)$ *and the parameter* $\kappa(\theta)$ *converge to finite limits* $\mathcal{E}_C$*,* $\mathcal{E}_Y$*,* $\mathcal{E}_{CY}$*,* $\mathcal{E}_S$ *and* κ as $\theta \to \overline{\theta}$*, with* $\mathcal{E}_S > 0$.

Under Assumption 1, we obtain

$$\lim_{\theta \to \overline{\theta}} \frac{B_C(\theta)}{H(\theta)} = \left[1 - \frac{\mathcal{E}_C}{\xi_C} + (1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}\right]^{-1}$$

and

$$\lim_{\theta \to \overline{\theta}} \frac{B_Y(\theta)}{H(\theta)} = \left[1 + \frac{\mathcal{E}_Y}{\xi_Y} + (1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}\right]^{-1}.$$

The representation of B_C requires that $1 + (1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y} > \frac{\mathcal{E}_C}{\xi_C}$; if this condition is violated then B_C is infinite, and thus the allocation cannot be optimal. B_Y on the other hand is finite and bounded above by 1. Similarly, we get

$$\lim_{\theta \to \overline{\theta}} \frac{B_S(\theta)}{H(\theta)} = \left[1 - \frac{\mathcal{E}_S}{\xi_S}\right]^{-1}$$

if $\mathcal{E}_S/\xi_S < 1$. If $\mathcal{E}_S/\xi_S \geq 1$, the allocation cannot be optimal, since in that case B_S is infinite, which implies that the marginal benefits to redistributing savings are infinitely large; but that would be inconsistent with the upper Pareto tail assumption. These expressions yield the following characterization of the optimal top income and savings taxes.

Theorem 1. *Suppose that the optimal allocation satisfies Assumption* 1*. Then the optimal labor wedge on top income earners satisfies*

$$\lim_{\theta \to \overline{\theta}} \tau_Y(\theta) = \tau_Y\left(\overline{\theta}\right) = \frac{\mathcal{E}_C/\xi_C + \mathcal{E}_Y/\xi_Y}{1 + \mathcal{E}_Y/\xi_Y + (1-\kappa)\mathcal{E}_{CY}/\xi_Y}, \tag{4}$$

and the optimal savings wedge on top income earners satisfies

$$\lim_{\theta \to \overline{\theta}} \tau_S(\theta) = \tau_S\left(\overline{\theta}\right) = \frac{\mathcal{E}_S/\xi_S + (1-\kappa)\mathcal{E}_{CY}/\xi_Y - \mathcal{E}_C/\xi_C}{1 - \mathcal{E}_S/\xi_S}, \tag{5}$$

where $1 + (1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y} > \frac{\mathcal{E}_C}{\xi_C}$ *and* $1 > \frac{\mathcal{E}_S}{\xi_S}$.

Equation (4) provides an extremely simple generalization of the standard top income tax rate formula of Saez (2001) to a dynamic environment. Equation (5) provides a sufficient statistics formula for savings taxes, analogous to that for income taxes, based on the ratio of marginal benefits of redistribution through consumption and savings. This principle can be extended to any number of commodities: the optimal wedge between any two goods is given by the ratio of marginal benefits of redistribution in the upper tail of the distribution. The parameter restrictions $1+(1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y} > \frac{\mathcal{E}_C}{\xi_C}$ and $1 > \frac{\mathcal{E}_S}{\xi_S}$ guarantee that the marginal benefits of redistributing consumption and savings are finite; otherwise the allocation cannot be optimal. They are imposed jointly on the primitive preference parameters and on the Pareto tails of the income, consumption, and savings distributions. They are, in principle, testable.

Upon controling for the shape of the distributions of income and consumption and the behavioral elasticities, the other primitives only enter through the determination of incentive-adjusted welfare weights. But these primitives disappear from the calculation of top income taxes whenever the highest types have incentive-adjusted welfare weights or marginal utilities that converge to zero. This last condition is implied by the Inada conditions on consumption and savings at the top of the type distribution. The result therefore holds without any further assumptions on primitives or observables.

In other words, Theorem 1 shows that it suffices to know eight statistics to compute the optimal top tax rates on labor and capital: three Pareto tail coefficients ξ_C, ξ_Y, ξ_S, three elasticities that govern the earnings (or labor supply) and consumption responses to income taxes $\mathcal{E}_C, \mathcal{E}_Y, \mathcal{E}_S$, the elasticity $\mathcal{E}_{CY}$ that measures preference complementarities between consumption and earnings, and the scaling parameter κ that captures the evolution of consumption along the earnings distribution. Importantly, we argue below that the model imposes overidentifying restrictions on these parameters.

Formula (4) shows that two terms change relative to the expression obtained by Saez (2001). The term $\mathcal{E}_C/\xi_C$ measures the impact of wealth effects on labor supply using the Pareto tail coefficient on consumption rather than income. It arises from the marginal benefits of redistributing consumption, and intuitively captures the notion that the marginal gains of further redistribution are linked to the tail of the consumption distribution. This term is present in Saez (2001) but with $\xi_C = \xi_Y$, since the static model ties consumption to after-tax income, which implies that the two must have the same Pareto tail. The term $(1-\kappa)\,\mathcal{E}_{CY}/\xi_Y$ adjusts the optimal income tax to control for substitution between consumption and leisure. This term is specific to the dynamic model we study here: it is equal to zero in a static economy where consumption coincides with after-tax income, i.e., where $\kappa = 1$.

The central observation here is that in dynamic economies, the optimal design of taxes should rely not only on income, but also on consumption data. In a static economy, consumption is equal to after-tax income and therefore does not provide any information not already contained in income data. This is no longer true once a consumption-savings margin is added to the model. In that case, consumption data is required to independently identify the wealth effects on labor supply that are captured by $\mathcal{E}_C/\xi_C$, and consumption and income data are required to identify the limit value of κ that is required to infer the complementarity term $(1-\kappa)\,\mathcal{E}_{CY}/\xi_Y$.

Alternatively, recall the equivalent interpretation of $1-\tau_Y$ as equalizing the marginal benefits of redistribution. Here it is intuitive that consumption data offers a more precise estimate of the marginal benefits of redistribution through consumption B_C, since these marginal benefits relate to how much the optimal tax system already manages to redistribute. And as discussed above, in a dynamic economy the marginal benefits of redistribution through earnings B_Y require an adjustment for complementarity between consumption and earnings that again requires a combination of income and consumption data.

4 Relationship with Saez (2001) and Quantitative Implications

In this section we discuss the relationship between the canonical optimal top tax rate formula obtained in a static setting by Saez (2001), and our generalized formulas (4) and (5) obtained by replacing the intra-temporal budget constraint with an inter-temporal one. We show that it is critical to distinguish three possible cases, based on the relative tail behavior of the income, consumption, and saving distributions.

4.1 Labor Supply and Consumption Responses to Taxes

As a preliminary step, we discuss how to transform optimal labor and savings wedges into tax schedules, and derive a correspondence between our elasticity parameters and the labor supply and consumption elasticities typically used in the literature.

From Wedges to Taxes. We focus on an implementation with a separate tax schedule $T_Y(Y)$ on income and $T_S(S)$ on savings. Given such tax schedules, the agent chooses (C, Y, S) to maximize utility $U(C, Y; \theta) + \beta V(S)$ subject to the budget constraint

$$S + T_S(S) \leq R\,(Y - T_Y(Y) - C)\,.$$

A tax schedule $\{T_Y(Y), T_S(S)\}$ implements an allocation $\{C(\theta), Y(\theta), S(\theta)\}$ if and only if the latter maximizes utility subject to the budget constraint, for all types. In the Appendix, we show that, under mild assumptions, these labor and savings taxes indeed implement the optimal labor and savings wedges if $T_Y'(Y(\theta)) = \tau_Y(\theta)$ and $T_S'(S(\theta)) = \tau_S(\theta)$.

Mapping Elasticities to Earnings and Consumption Responses. We suppose throughout that individual's earnings $Y(\theta)$ and consumption $C(\theta)$ are observed, and that savings are recovered from the inter-temporal budget constraint. Following the methodology of, e.g., Gruber and Saez (2002), we can use tax reforms to estimate income and substitution effects on earnings and consumption. These, in turn, map into the parameters of our analysis. Consider a reform that perturbs agent θ's total labor income tax liability by dT_Y, and the marginal labor income tax rate by dT_Y'. The earnings and consumption adjustments $dY(\theta)$, $dC(\theta)$ in response to such a reform can be expressed as

$$\begin{aligned} \frac{dY}{Y} &= -\zeta_Y^{sub}\frac{dT_Y'(Y)}{1-T_Y'(Y)} + \zeta_Y^{inc}\frac{dT_Y(Y)}{\left(1-T_Y'(Y)\right)Y} \\ \frac{dC}{C} &= -\zeta_C^{sub}\frac{dT_Y'(Y)}{1-T_Y'(Y)} - \zeta_C^{inc}\frac{dT_Y(Y)}{\left(1-T_Y'(Y)\right)Y}. \end{aligned} \tag{6}$$

The map between the elasticities $\left(\zeta_Y^{sub}, \zeta_Y^{inc}, \zeta_C^{sub}, \zeta_C^{inc}\right)$ and $(\mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S)$ is then given by:

$$\mathcal{E}_Y = \frac{1}{\zeta_Y^{sub} + \frac{\zeta_Y^{inc}\zeta_C^{sub}}{\zeta_C^{inc}}}, \quad \mathcal{E}_C = \frac{\zeta_Y^{inc}}{\zeta_C^{inc}}\mathcal{E}_Y, \quad \mathcal{E}_{CY} = \frac{1-\zeta_Y^{inc} - \zeta_Y^{sub}\mathcal{E}_Y}{\zeta_Y^{sub} - \frac{\gamma_C}{\pi_Y}\zeta_C^{sub}}, \quad \mathcal{E}_S = \frac{\frac{\pi_S}{\pi_Y}\frac{1}{R}\gamma_S\zeta_Y^{inc}}{\zeta_Y^{sub} - \frac{\gamma_C}{\pi_Y}\zeta_C^{sub}},$$

where $\gamma_C \equiv \frac{C}{Y-T_Y(Y)}$ and $\gamma_S \equiv \frac{S+T_S(S)}{Y-T_Y(Y)}$ denote respectively the consumption and savings shares of after-tax income, and where $\pi_Y \equiv \frac{1-T_Y'(Y)}{1-T_Y(Y)/Y}$ and $\pi_S \equiv \frac{1+T_S'(S)}{1+T_S(S)/S}$ are measures of the progressivity of the income and savings tax schedules. Note that $\gamma_C + \frac{1}{R}\gamma_S = 1$ and π_S, π_Y both converge to 1 if top marginal income and savings taxes converge to constants. The converse formulas, which give $\zeta_Y^{sub}, \zeta_Y^{inc}, \zeta_C^{sub}, \zeta_C^{inc}$ as functions of $\mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S$, are given in the Appendix.

4.2 Three Possible Scenarios

We show in the Appendix that the budget constraint imposes that $\xi_Y = \min\{\xi_C, \xi_S\}$. As a result, only three scenarios are possible, based on the relative values of the Pareto coefficients of the income, consumption and savings distributions: (i) $\xi_S > \xi_C = \xi_Y$, (ii) $\xi_C > \xi_S = \xi_Y$, and (iii) $\xi_C = \xi_S = \xi_Y$. Moreover, the consumption share converges to $\lim_{\theta\to\bar{\theta}}\gamma_C = 1$ in case (i), to $\lim_{\theta\to\bar{\theta}}\gamma_C = 0$ in case (ii), while in case (iii), $\lim_{\theta\to\bar{\theta}}\gamma_C$ can take on any value between 0 and 1.[8] Note finally that the

[8] Conversely, if $\lim_{\theta\to\bar{\theta}}\gamma_C \in (0,1)$, then $\xi_C = \xi_S = \xi_Y$. If $\lim_{\theta\to\bar{\theta}}\gamma_C = 1$, then $\xi_S \geq \xi_C = \xi_Y$. If $\lim_{\theta\to\bar{\theta}}\gamma_C = 0$, then $\xi_C \geq \xi_S = \xi_Y$.

consumption share and the Pareto coefficients discipline the parameter κ that appears in the optimal tax formulas via $\kappa = \frac{\xi_Y}{\xi_C}\gamma_C$.

In particular, this result implies that one cannot choose all three Pareto coefficients freely from the data. This is the analogue of the condition that $\xi_Y = \xi_C$ in the static setting, since the static budget constraint imposes that consumption and after-tax income are equal. Previewing our quantitative results, Toda and Walsh (2015) present empirical evidence that $\xi_C > \xi_Y$, which in turn imposes necessarily that $\xi_S = \xi_Y$. Conversely, variations of our model that would be consistent with higher wealth inequality than income inequality, such as Scheuer and Slemrod (2021), implicitly require that consumption must be as unequally distributed as savings, $\xi_C = \xi_S < \xi_Y$.[9] We return to this important point Section 4.4.

Case (i): Savings have a Thinner Tail than Income and Consumption

Suppose first that savings have a thinner tail than income ($\xi_S > \xi_C = \xi_Y$), and accordingly, the budget share of consumption γ_C converges to 1 for top earners (hence, $\frac{1}{R}\gamma_S \to 0$ and $\kappa \to 1$). In this case, we obtain $(1-\kappa)\,\mathcal{E}_{CY} = 0$, and our top income tax rate formula (4) reduces to

$$\tau_Y\left(\bar{\theta}\right) = \frac{\mathcal{E}_C/\xi_Y + \mathcal{E}_Y/\xi_Y}{1 + \mathcal{E}_Y/\xi_Y}.$$

On the other hand, the optimal tax formula derived in the static setting by Saez (2001, equation (8)) is given by:

$$\bar{\tau}_Y^{Saez} = \frac{1}{1 + \xi_Y \zeta_Y^{sub} - \zeta_Y^{inc}}. \tag{7}$$

Using the correspondence between $\zeta_Y^{sub}, \zeta_Y^{inc}$ and $\mathcal{E}_Y, \mathcal{E}_C$ derived in the Appendix with $\gamma_S = 0$, we get $\zeta_Y^{sub} = \frac{1/\mathcal{E}_Y}{1+\mathcal{E}_C/\mathcal{E}_Y}$ and $\zeta_Y^{inc} = \frac{\mathcal{E}_C/\mathcal{E}_Y}{1+\mathcal{E}_C/\mathcal{E}_Y}$. Substituting into equation (7) shows that $\bar{\tau}_Y^{Saez}$ coincides with $\tau_Y\left(\bar{\theta}\right)$. Thus, the static analysis delivers the correct optimal tax rate on labor income, and data on consumption (or savings) is not required to evaluate it. Intuitively, the dynamic model is equivalent to a static model at the top as the savings share of income converges to zero. Unfortunately, as we argue below, this case is not the empirically relevant one.

[9] Picking the three Pareto parameters ξ_Y, ξ_C, ξ_S freely from the data would require introducing an additional source of heterogeneity, which can be rates of return or endowments. While certainly this may be empirically plausible, incorporating such heterogeneity leads to complex multidimensional screening issues that the literature has not yet been able to fully address; for recent explorations of these questions, see e.g. Rothschild and Scheuer (2014) and Spiritus, Lehmann, Renes, and Zoutman (2021).

Case (ii): Consumption has a Thinner Tail than Income and Savings

Consider now the case where consumption has a thinner tail than income ($\xi_C > \xi_S = \xi_Y$), and the budget share of consumption converges to zero at the top, that is, $\gamma_C \to 0$ (and hence $\frac{1}{R}\gamma_S \to 1$ and $\kappa \to 0$). As we argue below, this case is likely to be the empirically relevant one.

Using the correspondence between $\zeta_Y^{sub}, \zeta_Y^{inc}$ and $\mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S$ with $\gamma_C = 0$, we get $\zeta_Y^{sub} = \frac{1/\mathcal{E}_S}{1+\mathcal{E}_{CY}/\mathcal{E}_S+\mathcal{E}_Y/\mathcal{E}_S}$ and $\zeta_Y^{inc} = \frac{1}{1+\mathcal{E}_{CY}/\mathcal{E}_S+\mathcal{E}_Y/\mathcal{E}_S}$. Substituting these expressions into equation (7) implies that $\bar{\tau}_Y^{Saez}$ can be equivalently expressed as

$$\bar{\tau}_Y^{Saez} = \frac{\mathcal{E}_S/\xi_Y + \mathcal{E}_{CY}/\xi_Y + \mathcal{E}_Y/\xi_Y}{1 + \mathcal{E}_Y/\xi_Y + \mathcal{E}_{CY}/\xi_Y}.$$

By contrast, our top labor and capital tax rates are given by (4) and (5), with ξ_S set equal to ξ_Y. It is clear that $\bar{\tau}_Y^{Saez}$ does *not* coincide with the optimum top labor income tax rate $\tau_Y\left(\overline{\theta}\right)$ obtained in our dynamic model, unless the relationship $\mathcal{E}_S/\xi_Y + \mathcal{E}_{CY}/\xi_Y = \mathcal{E}_C/\xi_C$ holds; that is, unless the optimum savings tax rate $\tau_S\left(\overline{\theta}\right)$ is equal to zero. More generally, the previous expressions lead to:

$$1 - \bar{\tau}_Y^{Saez} = \frac{1 - \tau_Y\left(\overline{\theta}\right)}{1 + \tau_S\left(\overline{\theta}\right)}.$$

This result implies that the static optimum $\bar{\tau}_Y^{Saez}$ overstates the correct optimum $\tau_Y\left(\overline{\theta}\right)$ whenever the optimal savings tax rate $\tau_S\left(\overline{\theta}\right)$ is strictly positive. Conversely, the static framework underestimates the optimum top labor income tax rate if savings are subsidized.

Intuitively, when the consumption share of top earners γ_C converges to zero, the optimal allocation of top earners is determined by a static trade-off between the two variables Y and S. Thus, the static optimum $\bar{\tau}_Y^{Saez}$ characterizes the optimal wedge between earnings and future consumption, which is a combination of the labor and savings wedges $\tau_Y\left(\overline{\theta}\right)$ and $\tau_S\left(\overline{\theta}\right)$. It follows that $\bar{\tau}_Y^{Saez}$ coincides with the optimal labor income tax rate $\tau_Y\left(\overline{\theta}\right)$ if and only if the optimal capital tax rate $\tau_S\left(\overline{\theta}\right)$ is equal to zero. But we saw in Corollary 2 that this is generally not the case: the optimal tax on savings is strictly positive whenever preferences are such that $U_{C\theta} < 0$, in which case the optimal tax on labor income is strictly lower than that predicted by the formula of Saez (2001).

The fact that the optimal labor tax rate is smaller (resp., larger) than the static optimum if capital is taxed (resp., subsidized) can also be understood by considering a marginal reduction in the labor income tax rate, starting from the static optimum. In addition to the standard effects on labor supply, this tax reform now also raises savings, which in turn raises (resp., lowers) government revenue if savings are taxed (resp., subsidized). The static model fails to account for this fiscal externality, hence overstates (resp., understates) the optimal tax rate on labor income.

Case (iii): Income, Consumption and Savings have Identical Tails

Finally, consider the case where the distributions of earnings, consumption, and savings all have the same tail coefficient ($\xi_Y = \xi_C = \xi_S$), and the budget shares of consumption γ_C and savings γ_S of top earners (and hence κ) converge to values strictly between 0 and 1.

Theorem 1 implies that the static optimum (7) generally differs from the optimal top tax rate formula (4). The dynamic adjustments can only be neglected when the first-period utility is quasi-linear in consumption, so that $U_{CC} = U_{CY} = 0$. Indeed, we have in this case $\mathcal{E}_C = \mathcal{E}_{CY} = \zeta_Y^{inc} = 0$ and $\mathcal{E}_Y = 1/\zeta_Y^{sub}$, so that the optimal labor income tax rate is equal to $\frac{\mathcal{E}_Y/\xi_Y}{1+\mathcal{E}_Y/\xi_Y}$ both in the static and the dynamic settings. However, whenever the utility of consumption has some curvature, even if preferences are GHH, the response of savings to labor income taxes modifies the optimal top tax rate on labor income, and the standard formula of Saez (2001) ceases to apply.

4.3 Calibration and Quantitative Implications

To determine which of the scenarios identified in the previous paragraph is empirically relevant, we can either compare the relative degrees of consumption and income inequality at the top, or estimate the consumption share and MPC of high income earners. These two approaches lead to different conclusions, and there is a fair amount of uncertainty regarding the values of the relevant parameters in the data; we thus propose two calibrations, one for case (ii) and one for case (iii) above.

4.3.1 Calibration for Case (ii)

Distribution Parameters ξ_Y, ξ_C, κ. The fact that the income distribution has a Pareto tail is well documented. In the U.S., the Pareto coefficient ξ_Y is approximately equal to 1.5 (P. Diamond and Saez (2011)), and certainly no larger than 2. Turning to the measures of consumption inequality at the top, Toda and Walsh (2015) argue that a Pareto tail fits the empirical distribution better than a log-normal, and they estimate an upper tail coefficient of $\xi_C = 3.65$. Thus, the distribution of consumption has a much thinner tail than that of pre-tax incomes. This suggests that the second case above, with $\gamma_C = \kappa = 0$ and $\frac{1}{R}\gamma_S = 1$, is the empirically relevant one: top earners save most of their income. Note that this result implies that the marginal propensity to consume converges to zero for high income earners, as equation (6) leads to $\frac{dC}{dT_Y} = \frac{\gamma_C}{\pi_Y}\zeta_C^{inc} = 0$.

Elasticities $\mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S$. There is a vast literature that estimates the elasticities of labor income with respect to marginal and average tax rates. Estimates of the Hicksian elasticity of taxable income ζ_Y^{sub} range between 0.1 and 0.5. The meta-analysis of Chetty (2012) leads to a preferred estimate of 0.33. For top income earners, Gruber and Saez (2002) estimate a value of 0.5. We evaluate optimal taxes for both values $\zeta_Y^{sub} \in \{0.33, 0.5\}$. Empirical evidence about the size of the income effects is mixed; see, e.g., Keane (2011). Several papers, for instance Gruber and Saez (2002), find small income effects, but others find larger values: e.g., Golosov, Graber, Mogstad, and Novgorodsky (2021) estimate that \$1 of additional unearned income reduces the pre-tax earnings of earners in the highest income quartile by 67 cents, which for a top marginal tax rate of 50 percent translates into an income effect of 0.33. We take a mid-range estimate of $\zeta_Y^{inc} = 0.25$. Note that the map between the two sets of elasticities derived above with $\gamma_C = 0$ pins down the values of $\mathcal{E}_Y + \mathcal{E}_{CY} = \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} \in \{2.25, 1.5\}$ and $\mathcal{E}_S = \frac{\zeta_Y^{inc}}{\zeta_Y^{sub}} \in \{0.75, 0.5\}$. Importantly, note that our calibration uses the estimates of labor supply responses to taxes $\zeta_Y^{sub}, \zeta_Y^{inc}$ and the Pareto coefficient $\xi_Y = \xi_S$ to pin down the combined wedge B_Y/B_S. Information about consumption only affects the breakdown between the two.

Direct empirical evidence about the elasticity parameters $\mathcal{E}_Y, \mathcal{E}_{CY}$ and $\mathcal{E}_C$ is scarce. As our baseline, we consider a model in which preferences are separable between C and Y ($U_{CY} = 0$). This in turn implies that $\mathcal{E}_{CY} = 0$ and $\mathcal{E}_C = \frac{-CU_{CC}}{U_C}$. We thus obtain $1/\mathcal{E}_Y \in \{0.44, 0.66\}$; note that in this case, $1/\mathcal{E}_Y$ is the Frisch elasticity of labor supply. Moreover $\mathcal{E}_C$ is the coefficient of relative risk aversion for consumption of top income earners; we choose three possible values $\mathcal{E}_C \in \{0.5, 1, 1.5\}$.[10]

In a second step, we allow for substitutability between consumption and leisure ($U_{CY} > 0$ and $\mathcal{E}_{CY} > 0$). With $\gamma_C = 0$, we still have $\mathcal{E}_C = -\frac{Cu_{CC}}{u_C}$.[11] We then follow Chetty (2006) to calibrate $\mathcal{E}_{CY}$, and thus obtain $\mathcal{E}_Y$. Chetty shows that the complementarity between consumption and labor can be bounded as a function of the coefficient of risk aversion by:

$$\mathcal{E}_{CY} = \frac{YU_{CY}}{U_C} \leq \frac{-CU_{CC}}{U_C} \times \frac{\Delta C/C}{\Delta Y/Y} = \mathcal{E}_C \times \frac{\Delta C/C}{\Delta Y/Y},$$

where $\frac{\Delta C/C}{\Delta Y/Y}$ is the change in consumption that results from an exogenous variation in labor supply (e.g., due to job loss or disability). He then estimates the latter parameter in the data and finds $\frac{\Delta C/C}{\Delta Y/Y} < 0.15$. We thus use the values $\mathcal{E}_{CY} \in \{0,\ 0.075\,\mathcal{E}_C,\ 0.15\,\mathcal{E}_C\}$ in our calibration.

[10] Assuming different functional forms for the utility function would yield different results. For instance, if the utility U is CES between consumption and leisure $1 - Y/\theta$, then $\mathcal{E}_C$ is the inverse of the elasticity of substitution. Aguiar, Hurst, and Karabarbounis (2012) give a preferred estimate of 2 for this elasticity; thus, we would get $\mathcal{E}_C = 0.5$. Obtaining credible empirical estimates of $\zeta_C^{sub}, \zeta_C^{inc}$ for top income earners by jointly estimating the system (6) is an important avenue for future research.

[11] This follows because $\frac{CU_{CY}}{-U_Y} = \gamma_C \frac{YU_{CY}}{U_C}$ and we are in the limit case where $\gamma_C \to 0$.

Optimal Top Tax Rates $\tau_Y\left(\overline{\theta}\right), \tau_S\left(\overline{\theta}\right)$ with Separable Preferences. Table 1 below summarizes our quantitative results for the optimal top tax rates on labor income and savings (the table reports the savings tax as a fraction of gross savings, $\tau_S\left(\overline{\theta}\right) / \left(1+\tau_S\left(\overline{\theta}\right)\right)$), for the case with separable preferences ($\mathcal{E}_{CY}=0$). In the static setting, formula (7) leads to $\bar{\tau}_Y^{Saez}=80\%$ for the low value of the Hicksian labor supply elasticity $\zeta_Y^{sub}=0.33$, and $\bar{\tau}_Y^{Saez}=67\%$ for the high value $\zeta_Y^{sub}=0.5$.

In the dynamic setting, the breakdown between savings and income taxes depends critically on the value of the elasticity $\mathcal{E}_C$. Recall that we are not imposing any *a priori* restriction on preferences, besides (in this paragraph only) the separability $U_{CY}=0$. Rather our calibration of the elasticities and Pareto tails implicitly determines the underlying structure of preferences. Thus, some parameter values can only be generated by $U_{C\theta}<0$, so that savings should be taxed, while others are only consistent with $U_{C\theta}>0$, so that savings should be subsidized.[12] For low values of the risk aversion ($\mathcal{E}_C=0.5$), the savings tax is high and the top labor income tax rate is substantially lower than its value in the static framework. For higher values of the risk aversion parameter $\mathcal{E}_C\in\{1,1.5\}$, the savings tax is lower and the income tax grows closer to the static optimum. For large enough values of $\mathcal{E}_C$, the optimum $\tau_Y\left(\overline{\theta}\right)$ exceeds $\bar{\tau}_Y^{Saez}$, in which case savings are subsidized, $\tau_S\left(\overline{\theta}\right)<0$.

To interpret the values of the savings wedge, it is useful to translate them into a tax on annualized returns. In our model, the first period represents a 30-year gap between the beginning of the working period and retirement. If the annual return on savings is 5%, a savings tax of $\frac{\tau_S}{1+\tau_S}=40\%$ corresponds to a 1.8% annual tax on accumulated wealth, or a 35% capital income tax. If annual returns are lowered to 3%, the implied tax on wealth is 1.7% and the implied capital income tax reaches 58%. These estimates are in the same ballpark as existing policy proposals of annual wealth taxes in the range of 1% to 2% (Saez and Zucman (2019b) and Saez and Zucman (2019a)). They also suggest that capital income should still be taxed at a significantly lower rate than labor income.

Alternatively, we can interpret our optimal tax system as a combination of income taxes, social security contributions and pension payments ("savings") that are indexed to labor income, without any additional private savings. The savings wedge then represents the marginal shortfall or excess of social security contributions relative to pension payments: a savings wedge of 40% at the top means that top income earners can only expect to receive a present value of 0.71 dollars of additional pension payments for each additional dollar in social security contributions.

To understand these results, it is useful to recall the characterizations of the marginal benefits of

[12] We revisit this discussion in Section 4.4 below.

Table 1: Case (ii), $\mathcal{E}_{CY} = 0$

	$\zeta_Y^{sub} = 0.33$		$\zeta_Y^{sub} = 0.5$	
	Income Tax τ_Y	**Savings Tax** $\frac{\tau_S}{1+\tau_S}$	**Income Tax** τ_Y	**Savings Tax** $\frac{\tau_S}{1+\tau_S}$
$\mathcal{E}_C = 0.5$	65%	42%	57%	22%
$\mathcal{E}_C = 1.0$	71%	31%	64%	8%
$\mathcal{E}_C = 1.5$	76%	15%	71%	−14%
Static Optimum $\bar{\tau}_Y^{Saez}$	80%		67%	

redistribution B_C, B_Y, and B_S given in Section 3. If income has a fatter Pareto tail than consumption ($\xi_C > \xi_S = \xi_Y$), the marginal benefits of redistributing earnings/leisure $B_Y(\theta)/H(\theta)$ converge to $(1 + \mathcal{E}_Y/\xi_Y + \mathcal{E}_{CY}/\xi_Y)^{-1}$, while the marginal benefits of redistributing savings $B_S(\theta)/H(\theta)$ converge to $(1 - \mathcal{E}_S/\xi_S)^{-1}$. The two elasticities in turn are pinned down uniquely by the income and substitution effects of earnings, ζ_Y^{inc} and ζ_Y^{sub}. Hence knowledge of these two income and substitution effects, along with an estimate of the common Pareto tail coefficient $\xi_S = \xi_Y$ is sufficient to estimate the combined labor and savings wedge

$$\frac{1 - \tau_Y\left(\overline{\theta}\right)}{1 + \tau_S\left(\overline{\theta}\right)} = \frac{B_Y(\theta)}{B_S(\theta)} = \frac{1 - \mathcal{E}_S/\xi_Y}{1 + \mathcal{E}_Y/\xi_Y + \mathcal{E}_{CY}/\xi_Y} = 1 - \bar{\tau}_Y^{Saez},$$

which yields the same value as the static optimum $\bar{\tau}_Y^{Saez}$ for the combined wedge. As discussed above, this is not surprising as the agents' decision problem for top income earners reduces to a static tradeoff between earnings and savings, for which the analysis of Saez (2001) continues to hold. With the calibrated parameters given above, we obtain $B_Y/H \in \{0.4, 0.5\}$, $B_S/H \in \{2, 1.5\}$, and $\bar{\tau}_Y^{Saez} \in \{0.8, 0.67\}$.

The other elasticity parameters then determine the marginal benefit of redistributing consumption B_C, which defines how the combined wedge $1 - \bar{\tau}_Y^{Saez}$ decomposes into separate labor and savings wedges. If $\mathcal{E}_S/\xi_S + \mathcal{E}_{CY}/\xi_Y \geq \mathcal{E}_C/\xi_C$, we have $B_S \geq B_C$, and it is optimal to tax savings; while if instead $\mathcal{E}_S/\xi_S + \mathcal{E}_{CY}/\xi_Y \leq \mathcal{E}_C/\xi_C$, then $B_S \leq B_C$, and it is optimal to subsidize them ($\tau_S\left(\overline{\theta}\right) \leq 0$). This breakdown then depends on the magnitude of the consumption elasticity, as well as the value of $\mathcal{E}_{CY}$, which captures the substitutability between consumption and leisure. Ignoring the latter by assuming that preferences are separable ($\mathcal{E}_{CY} = 0$), the optimal savings tax or subsidy then depends on the consumption elasticity of top income earners: the higher is $\mathcal{E}_C$, the higher is the tax on income and the lower is the tax on savings, until eventually the latter turns into a subsidy.

With separable preferences, our model also provides a lower bound on optimal income taxes and an upper bound on savings wedges. Since $B_C/H \geq 1$, we have $\tau_Y\left(\bar{\theta}\right) \geq 1 - B_Y/H \in \{0.6, 0.5\}$ and $\tau_S\left(\bar{\theta}\right) \leq B_S/H - 1 \in \{1, 0.5\}$. Hence, our model suggests a conservative lower bound on the optimal top income tax rate of 50% to 60%, and an upper bound on the optimal top savings tax between 33% and 50%. As Table 1 shows, the optimal top income and savings taxes span a considerable degree of variation between this lower bound when $\mathcal{E}_C$ is low, and the static optimum $\bar{\tau}_Y^{Saez}$ or even beyond, as $\mathcal{E}_C$ increases.

What is the rationale behind these numbers? The implication that static optimal labor taxes are very large results from a combination of low labor supply elasticities, which limit the efficiency cost of redistribution (or equivalently the marginal benefits of redistributing leisure), as well as a high degree of top income inequality and sufficiently high risk aversion to suggest that there are large gains from redistribution. The same logic implies a high combined wedge in the present environment. However, what determines whether this combined wedge should result in high taxes on earnings, or high taxes on savings? If consumption and savings elasticities are the same, then the fact that consumption appears to have a thinner tail than savings and top income earners save most of their income suggests that marginal benefits of redistribution are higher for savings than for consumption – and thus that it is optimal to load tax distortions into savings, rather than consumption, resulting in a lower income and a higher savings tax. Which of these marginal benefits dominates is then, again, a matter of the elasticity estimates on consumption vs. savings, along with the tail coefficients of the consumption and savings distributions.

These calibration results illustrate that viewing optimal tax design as an arbitrage between alternative margins of redistribution, instead of incentives vs. redistribution, is not just semantic, but leads to genuinely new economic insights, as well as a novel perspective on the sufficient statistics required to estimate optimal income and savings taxes.

Substitutability between Consumption and Leisure. So far, we have assumed that the utility is separable between consumption and leisure, so that $\mathcal{E}_{CY} = 0$. We now focus on the impact of non-separability between C and Y for optimal taxes by varying $\mathcal{E}_{CY}$ from 0 to $0.15\,\mathcal{E}_C$, in line with the upper bound suggested by Chetty (2006). We focus on the case $\zeta_Y^{sub} = 0.33$ and $\zeta_Y^{inc} = 0.25$, so that $\mathcal{E}_Y + \mathcal{E}_{CY} = \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} = 2.25$ and $\mathcal{E}_S = \frac{\zeta_Y^{inc}}{\zeta_Y^{sub}} = 0.75$. These parameters imply $B_Y/H = 0.4$, $B_S/H = 2$, and $\bar{\tau}_Y^{Saez} = 0.8$. Substitutability between consumption and leisure thus leaves the combined labor and savings wedge unchanged but shifts the wedge from labor to savings taxes.

The resulting values for the optimal tax rates are given in Table 2. Note that the first two columns are the same as those of Table 1 by construction. We recover the familiar result that when preferences are non-separable, it is optimal to tax less heavily the goods that are complementary to

Table 2: Case (ii), $\mathcal{E}_{CY} \geq 0$

	$\mathcal{E}_{CY} = 0$		$\mathcal{E}_{CY} = 0.075\,\mathcal{E}_C$		$\mathcal{E}_{CY} = 0.15\,\mathcal{E}_C$	
	τ_Y	$\frac{\tau_S}{1+\tau_S}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$
$\mathcal{E}_C = 0.5$	65%	42%	65%	44%	63%	45%
$\mathcal{E}_C = 1.0$	71%	31%	69%	35%	67%	39%
$\mathcal{E}_C = 1.5$	76%	15%	73%	25%	70%	32%

labor (Corlett and Hague (1953)). In our setting, where earnings and consumption are complements, the planner reduces the tax rate on labor income and raise the tax rate on savings.

Quantitatively, the complementarity correction has a fairly small impact on the optimal labor and savings tax rates for reasonable empirical values of $\mathcal{E}_{CY}$. Formulas (4) and (5) imply that the correction for complementarity $\mathcal{E}_{CY}/\xi_Y$ is equivalent to adjusting the Pareto tail coefficient on consumption upwards to $\widetilde{\xi}_C$ defined by $\mathcal{E}_C/\widetilde{\xi}_C = \mathcal{E}_C/\xi_C - \mathcal{E}_{CY}/\xi_Y$, and the Pareto coefficient on earnings downwards to $\widetilde{\xi}_Y$ defined by $\mathcal{E}_Y/\widetilde{\xi}_Y = \mathcal{E}_Y/\xi_Y + \mathcal{E}_{CY}/\xi_Y$. It thus amounts to increasing the effective gap between income and consumption inequality. The adjustment increases the consumption tail coefficient from $\xi_C = 3.65$ to $\widetilde{\xi}_C = 5.75$. For $\mathcal{E}_C = 1$, this lowers the marginal benefit of redistributing consumption B_C/H from 1.38 to 1.21, equivalent to a 12% increase in after-tax labor income and a corresponding increase in the savings wedge. For $\mathcal{E}_C = 0.5$ and $\mathcal{E}_C = 1.5$, the corresponding values are respectively 5.5% and 20%.

The corresponding adjustment for the income tail coefficient depends on the chosen value of $\mathcal{E}_C$, but is generally much smaller, with an adjustment from $\xi_Y = 1.5$ to values of $\widetilde{\xi}_Y \in \{1.45, 1.41, 1.36\}$ if $\mathcal{E}_C \in \{0.5, 1, 1.5\}$. These adjustments barely change B_Y/H which decreases from 0.4 at the benchmark without complementarity to no lower than 0.38 when $\mathcal{E}_C = 1.5$. Finally B_S/H does not change with $\mathcal{E}_{CY}$.

In line with this observation that B_S is unchanged and B_Y barely affected, we observe that the combined wedge $\bar{\tau}_Y^{Saez}$ varies little, from 80% to no more than 83% when $\mathcal{E}_C = 1.5$. Therefore, most of the adjustment comes from the change in B_C, which shifts the tax distortion at the top from income to savings taxes. A ballpark estimate across these different cases is that $\mathcal{E}_{CY}$, at the bound suggested by Chetty, pushes labor income taxes between one third and half of the way towards their theoretical bound that is obtained by setting $\mathcal{E}_C = 0$.

Savings should be taxed if and only if $\frac{\mathcal{E}_S}{\mathcal{E}_C} > \frac{\xi_S}{\widetilde{\xi}_C}$ where $\widetilde{\xi}_C$ is the adjusted Pareto tail coefficient. Without the complementarity correction, we have $\xi_S = 1.5$, $\xi_C = 3.65$, and $\mathcal{E}_S = 0.75$, so that savings should be taxed unless the risk aversion coefficient $\mathcal{E}_C$ is larger than $\frac{\xi_C}{\xi_S}\mathcal{E}_S = 1.8$. With the complementarity correction, we have $\widetilde{\xi}_C = 5.75$, so risk aversion $\mathcal{E}_C$ needs to exceed 2.9 to overturn the conclusion that savings should be taxed. To sum up, already without complementarity the marginal benefit of redistributing savings appear to be high relative to the marginal benefit of redistributing consumption, and consumption has a much thinner upper tail than income and savings. The correction only reinforces this conclusion. So unless $\mathcal{E}_C$ is very large, the marginal benefits of redistributing consumption remain substantially smaller than the marginal benefits of redistributing savings, resulting in a significant shift from income to savings taxes at the optimal allocation.

4.3.2 Calibration for Case (*iii*)

The previous calibration assumed that the distribution of consumption has a thinner tail than that of incomes, $\xi_C > \xi_Y$. As we noted, this implies that the marginal propensity to consume out of labor income converges to zero for top earners. This may be counterfactual. The estimates of Auclert (2019) (Figure 2), for instance, suggest that the MPC decreases with income and converges to about 0.2, although this estimate is based on survey data where the top 1 percent earners are missing. On the other hand, Lewis, Melcangi, and Pilossoph (2019) find that high incomes tend to have higher MPCs. This would suggest that case (*iii*) may be the empirically relevant one. We thus propose another calibration to compute the optimal income and savings taxes in this case.

We proceed along similar lines as before in case (*ii*): We take $\zeta_Y^{sub} = 0.33$ and $\zeta_Y^{inc} = 0.25$ as in the previous calibration. As our benchmark, we assume that preferences are separable between C and Y, so that $\mathcal{E}_{CY} = 0$ and $\mathcal{E}_C = \frac{-CU_{CC}}{U_C}$. We choose the same values for the risk aversion coefficient as before but in addition also include 0.75, for reasons that we explain below. The expression for $\mathcal{E}_{CY}$ given by the representation of Section 4.1 implies that $\mathcal{E}_Y = \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}}$. This leads to $\mathcal{E}_Y = 2.25$. So far, these values are exactly the same as in the calibration of case (*ii*).

Recall that the Pareto coefficients of consumption, earnings, and savings must coincide, $\xi_Y = \xi_C = \xi_S$. We start by setting this parameter to 1.5, the value we used for income and savings in the calibration of case (*ii*), so as to make the results as closely comparable as possible. We then vary the Pareto coefficient to 2 (a mid-range between the coefficients estimated by Toda and Walsh (2015)), and 3.65 (the value of the tail coefficient for consumption).[13]

We are left with selecting $\mathcal{E}_S$ and the consumption share γ_C. Their values depend on the income and substitution effects of consumption ζ_C^{inc} and ζ_C^{sub}, for which we do not have readily available empirical counter-parts. However it follows from the representation of $\mathcal{E}_C$ and $\mathcal{E}_Y$ that

[13] Note that, in the corresponding optimum formula (4), one can use either the Pareto coefficient of incomes or consumption, since $\xi_Y = \xi_C$. To the extent that consumption is a better proxy than income for lifetime earnings, it may be preferable to use the estimate of the Pareto tail on consumption in the model.

$\mathcal{E}_C/\mathcal{E}_Y = \zeta_Y^{inc}/\zeta_C^{inc}$. In addition, $1/\mathcal{E}_Y = \zeta_Y^{sub} + \zeta_Y^{inc}\zeta_C^{sub}/\zeta_C^{inc}$ pins down the ratio $\zeta_C^{sub}/\zeta_C^{inc} = \mathcal{E}_Y^{-1}$. Hence, we can express ζ_C^{inc} and ζ_C^{sub} in terms of $\mathcal{E}_Y$, $\mathcal{E}_C$, and ζ_Y^{inc}. We calibrate γ_C to match an MPC of top income earners of $\frac{dC}{dT_Y} = 0.2$, which by equation (6) is equal to $\gamma_C\zeta_C^{inc}$. We obtain $\gamma_C = 0.356\,\mathcal{E}_C$ resulting in values for the budget share of consumption γ_C for top earners between 0.178 and 0.533 for the values of $\mathcal{E}_C$ considered above.

The last parameter, $\mathcal{E}_S$, is pinned down by the representation of Section 4.1, which implies $\mathcal{E}_S = \mathcal{E}_C\frac{1-\gamma_C}{\mathcal{E}_C\zeta_Y^{sub}/\zeta_Y^{inc}-\gamma_C}$. Consider first the case where $\mathcal{E}_C = \zeta_Y^{inc}/\zeta_Y^{sub} = 0.75$, which corresponds to the value that we have added to the current calibration. In this case, the above equation implies that $\mathcal{E}_S = \mathcal{E}_C = \zeta_Y^{inc}/\zeta_Y^{sub}$, i.e., the two elasticities are identical and pinned down by the ratio of income and substitution effects of earnings. When instead $\mathcal{E}_C \gtrless \zeta_Y^{inc}/\zeta_Y^{sub}$, it then follows that $\zeta_Y^{inc}/\zeta_Y^{sub} \gtrless \mathcal{E}_S$. In summary, $\mathcal{E}_S$ is decreasing with the choice of $\mathcal{E}_C$ and they cross each other exactly when $\mathcal{E}_S = \mathcal{E}_C = \zeta_Y^{inc}/\zeta_Y^{sub}$. With the target of $\gamma_C = 0.356\,\mathcal{E}_C$ and $\zeta_Y^{sub}/\zeta_Y^{inc} = 0.75$, the implied values for $\mathcal{E}_S$ vary from 0.841 when $\mathcal{E}_C = 0.5$ to 0.478 when $\mathcal{E}_C = 1.5$.

Table 3 gives the results, along with the combined wedge $\bar{\tau} \equiv 1 - \frac{1-\tau_Y}{1+\tau_S}$. First, when $\mathcal{E}_S = 0.75$, the savings wedge τ_S is equal to zero and the labor wedge τ_Y is equal to the static wedge $\bar{\tau}_Y^{Saez}$. This is a consequence of the Atkinson and Stiglitz (1976) theorem (Corollary 2). When $\mathcal{E}_{CY} = 0$ and $\mathcal{E}_S = \mathcal{E}_C$, preferences are separable and the utility of consumption is homogeneous across consumers. In this case, the benefits of redistributing via consumption and savings are then identical, since $B_S/H = (1 - \mathcal{E}_S/\xi_S)^{-1}$ and $B_C/H = (1 - \mathcal{E}_C/\xi_C + \mathcal{E}_{CY}/\xi_Y)^{-1}$. Away from this benchmark, it is optimal to distort savings even though preferences are still separable between consumption and earnings ($U_{CY} = 0$). Indeed, the utility of consumption is no longer homogeneous across consumers ($U_{C\theta} \neq 0$) when $\mathcal{E}_S \neq \mathcal{E}_C$. The savings wedge is positive, and the labor wedge smaller than the static wedge, whenever $\mathcal{E}_S > \mathcal{E}_C$. However the increase in the savings wedge more than offsets the decline in the labor wedge so that the combined wedge goes up. Conversely, when $\mathcal{E}_S < \mathcal{E}_C$, so that higher types have higher consumption needs ($U_{C\theta} > 0$), income bears more than 100% of the combined wedge $\bar{\tau}$, while savings are subsidized.[14] Note finally that the constraint $1 > \mathcal{E}_C/\xi_C$ binds when $\mathcal{E}_C = \xi = 1.5$; in this extreme case, our formulas imply that $\tau_Y = 1$ and $\tau_S = -1$.

We finally evaluate the impact of the complementarity between consumption and earnings on our quantitative results. As before we let $\zeta_Y^{sub} = 0.33$, $\zeta_Y^{inc} = 0.25$, and $\frac{dC}{dT_Y} = 0.2$. We assume that the Pareto coefficient is $\xi = 2$. The calibration procedure is more complicated in this case and is described in the Appendix. As in the calibration of Case (*ii*), we use Chetty (2006)'s empirical estimate of the complementarity $\mathcal{E}_{CY}$ as a fraction α of the coefficient of risk aversion

[14] Note that when $\gamma_C > 0$, the combined wedge $\bar{\tau}$ is in general different from the static wedge $\bar{\tau}^{Saez}$.

Table 3: Case (iii), $\mathcal{E}_{CY} = 0$

	$\xi = 1.5$			$\xi = 2$			$\xi = 3.65$		
	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$
$\mathcal{E}_C = 0.5$	73%	34%	82%	65%	22%	73%	47%	11%	53%
$\mathcal{E}_C = 0.75$	80%	0%	80%	71%	0%	71%	51%	0%	51%
$\mathcal{E}_C = 1.0$	87%	-67%	78%	76%	-33%	68%	55%	-12%	50%
$\mathcal{E}_C = 1.5$	N/A	N/A	N/A	88%	-203%	64%	64%	-47%	47%
$\bar{\tau}_Y^{Saez}$	80%			71%			51%		

Table 4: Case (iii), $\mathcal{E}_{CY} \geq 0$

	$\mathcal{E}_X = 0$			$\mathcal{E}_X = 0.1\,RRA$			$\mathcal{E}_X = 0.15\,RRA$		
	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$	τ_Y	$\frac{\tau_S}{1+\tau_S}$	$\overline{\tau}$
$\mathcal{E}_C = 0.5$	65%	22%	73%	64%	26%	73%	63%	28%	73%
$\mathcal{E}_C = 0.75$	71%	0%	71%	69%	6%	71%	69%	9%	72%
$\mathcal{E}_C = 1.0$	76%	-33%	68%	75%	-23%	69%	74%	-19%	69%
$\mathcal{E}_C = 1.5$	88%	-203%	64%	86%	-163%	63%	86%	-150%	65%

RRA. This procedure pins down both $\mathcal{E}_{CY}$ and the endogenous consumption share γ_C and, in turn, $\mathcal{E}_Y$ and $\mathcal{E}_S$. We take $\alpha = 0.15$, the highest value reported by Chetty, and an intermediate degree of complementarity $\alpha = 0.1$. The corresponding values of $\mathcal{E}_Y, \mathcal{E}_{CY}, \mathcal{E}_S, \gamma_C$ are reported in the Appendix.

Table 4 gives the results. By construction, the first two columns on the left are identical to the middle columns of Table 3. As in Case (ii), the complementarity between consumption and income raises the optimal savings wedge and lowers the labor wedge. For reasonable values of $\mathcal{E}_{CY}$, these adjustments are negligible for the income tax, but they may be sustantial for the savings tax: τ_S must be adjusted upwards by up to 30 percent if $\mathcal{E}_C = 0.5$.

4.4 Sufficient Statistics: A Note of Caution

Many papers in the literature impose strong *a priori* assumptions on the utility function to derive optimal taxes in terms of elasticity parameters and Pareto coefficients, before resorting to empirical estimates of these parameters to evaluate the formulas quantitatively. As emphasized by Chetty (2009), a potential pitfall of this "sufficient statistic" approach is that these empirical estimates may not be compatible with the structural restrictions imposed by the model. For instance, the optimal tax formula in the static Mirrlees-Saez setting relies on the strong assumption that consumption equals after-tax income – we saw that, if this is not the case, the Pareto coefficients of both income and consumption generally appear in the optimal top tax rate formula. Yet, we argued that this condition is rejected by the data. The literature typically ignores this issue by simply using empirical estimates of the Pareto coefficient of the income (rather than consumption) distribution in the optimal income tax formula.

What are the restrictions on observable parameters that our framework imposes? To answer this question, differentiate the individual's problem first order conditions $-U_Y/U_C = 1 - \tau_Y$ and $U_C = \frac{\beta R}{1+\tau_S} V'$ with respect to θ. We obtain the following condition:

$$\frac{\frac{\mathcal{E}_C}{\xi_C} - \frac{\mathcal{E}_S}{\xi_S} - (1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}}{\frac{\mathcal{E}_C}{\xi_C} + \frac{\mathcal{E}_Y}{\xi_Y}} = \lim_{\theta \to \bar{\theta}} \frac{\frac{U_{C\theta}}{U_C}}{\frac{U_{C\theta}}{U_C} - \frac{U_{Y\theta}}{U_Y}}. \tag{8}$$

Therefore, picking values for the empirically observable parameters $\xi_Y, \xi_C, \xi_S, \mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S, \kappa$ that appear on the left hand side of this equation pins down the value of the right hand side. In other words, the elasticity parameters and Pareto tails impose a weight on consumption- vs. needs-based redistribution motives. In our calibration, we did not impose any functional form assumptions on preferences. Thus, the empirical estimates we used are consistent with our underlying structural model. But the "data" implicitly tells us what is the underlying structure on the parameter in the right hand side of (8).

Conversely, one cannot make an *a priori* assumption on the shape of the utility function to derive an optimal tax formula, and then "let the data speak" about the elasticity and Pareto parameters – the resulting value of the optimal tax rate could then be inconsistent with the structural model that led to the formula. For instance, suppose that the values of the calibrated parameters imply that the left hand side of (8) is strictly negative, as is most often the case in our calibration. This overidentifying restriction is inconsistent with, e.g., separable preferences with a marginal utility of consumption that is independent of θ.[15] To take an even more striking example, suppose that optimal taxes were derived under the assumptions that preferences are GHH, $U = u\left(g\left(C\right) - v\left(Y/\theta\right)\right)$ for some convave constant elasticity functions u and g and convex function v. This is a typical assumption made in the literature. While this utility function implies $U_{C\theta} \leq 0$, which at first sight seems consistent with (8), the Inada condition implies that $\frac{v(Y(\theta)/\theta)}{g(C(\theta))}$ must converge to a positive constant as $\theta \to \bar{\theta}$. At the same time, $1 - \tau_Y = \frac{v'(Y(\theta)/\theta)/\theta}{g'(C(\theta))}$ also converges to a constant. These two conditions yield, respectively, $\frac{1+\mathcal{E}_C}{\xi_C} = \frac{1+\mathcal{E}_Y}{\xi_Y} - \lim_{\theta \to \bar{\theta}} H\left(\theta\right)/\theta$ and $\frac{\mathcal{E}_C}{\xi_C} = \frac{\mathcal{E}_Y}{\xi_Y} - \lim_{\theta \to \bar{\theta}} H\left(\theta\right)/\theta$ and therefore can both hold only if $\xi_C = \xi_Y$, i.e., if income and consumption have the same Pareto tail. On the other hand, if $\xi_C > \xi_Y$, the Inada condition still holds, but with $\frac{v(Y(\theta)/\theta)}{g(C(\theta))}$ and hence the right-hand

[15] Such preferences would impose $\mathcal{E}_C/\xi_C - \mathcal{E}_S/\xi_S = 0$ and $\mathcal{E}_{CY} = 0$, and hence $\tau_S\left(\bar{\theta}\right) = 0$, consistent with Corollary 2. That is, the ratio of elasticities is pinned down by the ratio of Pareto coefficients.

side of (8) converging to 0. As a result, this functional form must either violate the restriction (8), or impose that $\xi_C = \xi_Y$, which as we discussed is not consistent with empirical evidence.

5 Extensions

We conclude our paper by discussing several possible extensions. We argue that variants of the redistributional arbitrage formulas $1 - \tau_Y = \frac{B_Y}{B_C}$ and $1 + \tau_S = \frac{B_S}{B_C}$ remain applicable in richer environments, as well as for other margins of taxation, even if the characterization of marginal benefits of redistribution may change.

Optimal Tax Schedule and Bottom Income Taxes

Although the B_Y/B_C- and B_S/B_C-formulae of optimal income and savings wedges remain valid throughout the type space, our analysis so far only focused on implications of optimal taxes on top income earners. However, it is straight-forward to use the same analysis to draw additional implications throughout the tax schedule.

For example, differentiating the formula with respect to θ yields the following expression:

$$\frac{\tau_Y'(\theta)}{1 - \tau_Y(\theta)} = \frac{B_C'(\theta)}{B_C(\theta)} - \frac{B_Y'(\theta)}{B_Y(\theta)}$$

This expression equates the marginal change in the tax distortion, a measure of local tax progressivity, on the left hand side, to the growth rate of the marginal benefit of redistributing consumption minus growth rate of the marginal benefit of redistributing leisure, at any given type. Marginal taxes are thus progressive (increasing with type) if and only if the marginal benefit of redistributing consumption displays higher growth along the type distribution than the marginal benefit of redistributing leisure. Some tedious but relatively straight-forward derivations allow us to represent the right hand side as follows:

$$\frac{\tau_Y'(\theta)}{1 - \tau_Y(\theta)} \frac{Y(\theta)}{Y'(\theta)} = \xi_Y(\theta) \left(\frac{\mathcal{E}_C(\theta)}{\xi_C(\theta)} + \frac{\mathcal{E}_Y(\theta)}{\xi_Y(\theta)} - \left(\frac{1}{B_Y(\theta)} - \frac{1}{B_C(\theta)} \right) \right)$$

where $\xi_Y(\theta) \equiv \frac{Y'(\theta)}{Y(\theta)} H(\theta)$ and $\xi_C(\theta) \equiv \frac{C'(\theta)}{C(\theta)} H(\theta)$ represent generalized tail coefficients of the earnings and consumption distribution (these are equal to ξ_Y and ξ_C if the distributions are exactly Pareto). The left hand side measures local tax progressivity as the elasticity of the labor wedge $1 - \tau_Y$ to earnings Y. Optimal tax progressivity is thus expressed as a function of the preference elasticities $\mathcal{E}_C(\cdot)$ and $\mathcal{E}_Y(\cdot)$ that govern how the agents' marginal rate of substitution responds to marginal changes in consumption and earnings, the upper tail of the earnings and consumption distribution as expressed by $\xi_Y(\cdot)$ and $\xi_C(\cdot)$, and the gap between $1/B_Y$ and $1/B_C$ which governs the relative change of the planner's weights on consumption vs. leisure at the top of the earnings distribution. A similar characterization of optimal tax progressivity also obtains for the optimal savings wedge.

Likewise, by re-scaling the marginal benefits of redistribution by $\frac{1-F(\theta)}{F(\theta)}$, we can obtain analogues of the B_Y/B_C- and B_S/B_C-formulae for bottom income and savings taxes, which can then be mapped to the same preference elasticities along with the lower tail of income, savings and consumption distributions. Additional complications which take us beyond the scope of this paper arise because of an absence of a suitable analogue to the Inada condition. Nevertheless, the approach of optimal income taxes as redistributional arbitrage can also be applied to understanding tax distortions at the bottom of the income distribution. We leave a full analysis of income taxation and redistributional policies for low income earners for future work.

General Preferences

We have assumed that preferences were additively time-separable, so that the benefits of "savings" were independent of type θ, "consumption" and "earnings". The principal simplification brought about by time-additively separable preferences is that the marginal benefit of redistribution through savings B_S doesn't require its own incentive-adjusted probability measure. It is straight-forward to generalize our results to preferences of the form $U(C, S, Y; \theta)$, in which case a further change of probability measures appears in the computation of B_S.[16] The uniform commodity taxation principle then applies whenever at the optimal allocation, $\frac{U_{\theta C}}{U_C} = \frac{U_{\theta S}}{U_S}$ for all θ, i.e., marginal rates of substitution between consumption and savings are homogeneous across all agents.

In addition, the separability assumption imposes some structure on income and substitution effects of the different commodities. This in turn simplifies the identification of sufficient statistics. As we have seen above, the computation of the top income and savings taxes requires estimates of four preference parameters, three elasticities and an adjustment for complementarity between consumption and earnings. With unrestricted preferences, the analysis will require estimates for two additional preference elasticities to account for complementarity of consumption and earnings with savings.

[16] In addition, agents' first-order conditions then tie the marginal utility of consumption U_C to the marginal utility of savings, and the latter can be directly mapped to the observation of savings. For $\mathcal{E}_S > 0$, this allowed us to validate the Inada condition that the top types of the distribution have vanishing weight in the planner's objective, thus simplifying the characterization of top income taxes. Without additive separability, the Inada condition must be assumed as a property of the optimal allocation, or established by other arguments.

Alternative Social Objectives

Our analysis constructed the marginal benefits of redistribution under a utilitarian or expected utility welfare criterion. It is straight-forward, however, to extend our arguments to any alternative planner objective. The model introduced in Section 2 minimizes the planner's resource cost subject to incentive compatibility conditions and a "promise-keeping" constraint. We can consider alternative planner objectives which replace the utilitarian promise-keeping constraint $\mathbb{E}[W(\theta)] \geq v_0$, where $W(\cdot) \equiv U(C, S, Y; \theta)$ represents the agent's indirect utility as a function of type, with any alternative $\mathbb{E}[\omega(\theta)\mathcal{G}(W(\theta))] \geq \hat{v}_0$, where $\omega(\cdot) > 0$ corresponds to a type-specific modification of Pareto weights with $\mathbb{E}[\omega(\theta)] = 1$, and $\mathcal{G}(\cdot) > 0$ represents the planner's attitudes towards inequality, relative to the utilitarian benchmark.

Defining $\mathcal{U}(C, S, Y; \theta) \equiv \omega(\theta)\mathcal{G}(U(C, S, Y; \theta))$, we can then perform exactly the same analysis as in our paper with $\mathcal{U}(\cdot)$ instead of $U(\cdot)$. Importantly, this modification keeps incentive compatibility constraints, and hence the adjusted probability measures unchanged.[17] The alternation of social planner's objectives only enters into how the marginal benefits of redistribution are calculated: the required inverse marginal utilities are then based on a combination of the agents' and the planner's preferences.

Strikingly, one can then show that the top income and savings taxes do not depend on the planner's attitudes towards redistribution, i.e., they are independent of $\omega(\cdot)$ and $\mathcal{G}(\cdot)$ and only depend on $U(\cdot)$, as long as the Inada condition implies that top types receive vanishing welfare weights. Indeed, we have seen that under the Inada condition, marginal benefits of redistribution at the top are a function only of the incentive adjustments, which depend on $U(C, S, Y; \theta)$ but not on $\omega(\cdot)$ and $\mathcal{G}(\cdot)$. If the planner attributes zero welfare weight to the richest households, then the design of top tax rates comes down to an efficiency trade-off between asking these households to work more, vs. asking them to consume and save less, in order to maximize the resources that can be extracted at the top and redistributed towards lower types. This local efficiency tradeoff at the top of the type distribution is independent of the planner's global attitudes towards inequality.

Additional Commodities and Alternative Model Interpretations

Another possible direction for extension is to consider more than two commodities, or alternative interpretations of the commodity bundle. For example, we could relabel S in our model as "bequests", and let C and Y stand for life-time income and consumption. In this case our results would

[17] Formally, the transformation from $U(\cdot)$ to $\mathcal{U}(\cdot)$ does not alter the marginal rates of substitution between consumption, savings, and income, for any given type.

reinterpret the savings tax as a tax on bequests. Alternatively, we could interpret C as "basic necessities" and S as "luxury goods" in a static interpretation of our model. In this case the savings tax represents a relative price distortion between the two, possibly in the form of subsidies on basic necessities. In kind subsidies for basic necessities, such as subsidized rent, food stamps, public transportation, education or health services play a central role in increasing the welfare of low-income households. On the other hand, governments may also find it opportune to tax certain consumption goods favored by higher income households.

With a one-dimensional preference type, our analysis can directly be extended to multiple consumption goods. This leads to an immediate generalization of the result that optimal relative price distortions can be characterized as arbitraging between redistribution through one commodity vs. the other. In particular, following the same steps as above, one can show that the optimal wedge $\tau_{j,k}$ between two commodities j and k is characterized by

$$1 - \tau_{j,k} = \frac{B_j}{B_k}$$

where B_j and B_k represent marginal benefits of redistributing commodity j or k from types above θ to types below θ, and defined by the same incentive-adjusted probability measures as described in Section 2.

As we discussed in the context of Corollary 2, our model reveals a potential rationale for non-uniform commodity taxation for redistributive objectives, which our model displayed through savings taxes. This rationale arises whenever two different commodities yield different incentive-adjusted probabilities. Potential departures from uniform commodity taxation are then linked to these incentive-adjusted probability measures which in turn can be mapped to observables.

Our analysis thus develops a template for future empirical work that seeks to identify optimal commodity taxes and subsidies by identifying the required marginal benefits of redistribution for any commodity, using observed distributions of consumption and estimated demand elasticities. One key application of this framework may be to housing which is an important budget component of most households, thus displaying important wealth effects, and which benefits from a whole array of redistributive interventions, from subsidized public housing or rent subsidies at the low end of the income distribution to mortgage interest deductions at the upper end. Our analysis may offer an efficiency rationale for implementing such policies, as well as practical guidance on how such policies should be structured to achieve the government's redistributive objective.

Richer Dynamics

Another natural direction is to follow the lead of dynamic Mirrlees models (Golosov, Kocherlakota, and Tsyvinski (2003), Farhi and Werning (2013), and Golosov, Troshkin, and Tsyvinski (2016)) and allow for stochastic evolution of types over multiple periods. Hellwig (2021) develops the implications of such a model for labor and savings wedges with non-separable preferences, and derives conditions, linked to the persistence of types and information rents, under which the optimal labor wedge can still be represented as $1 - \tau_Y\left(\theta^t\right) = \frac{B_Y\left(\theta^t\right)}{B_C\left(\theta^t\right)}$, where θ^t represents a t-period sequence of type realizations.[18] The interpretation of this formula as a redistributional arbitrage is the same as here, but in contrast to the present static setting, the current marginal benefits are now based on distributions of earnings and consumption growth conditional on the prior sequence of types, or equivalently, the prior earnings history.

Multi-Dimensional Types

The assumption of a one-dimensional type space becomes more difficult to justify as one moves beyond a single consumption good since there is no reason why individual ability should be perfectly aligned with tastes for different commodities, for example. In line with this assumption our derivation of sufficient statistics made use of the fact that consumption, earnings, and savings were perfectly co-monotonic at the optimal solution. Such perfect co-monotonicity seems extremely implausible from an empirical point of view, even with a simple commodity space of three goods, like ours.

Another natural extension is therefore to extend the present analysis to multi-dimensional type spaces. While multi-dimensional screening is notoriously challenging, preliminary results in Hellwig (2022b) for a multi-good monopolist problem suggest that core ideas from the present analysis can be generalized, in particular the representation of local incentive compatibility through incentive-adjusted probability measures, the characterization of optimal relative price distortions through an arbitrage of information rents, and a general representation of optimal distortions that generalizes the B_j/B_k-formula presented here. On the other hand, the ABC-formula does not generalize, since the tradeoff between efficiency and redistribution (or information rents) is no longer well-defined for multi-dimensional type spaces.[19] These preliminary results suggest that there is scope to generalize

[18] When this optimality condition doesn't hold exactly, $\frac{B_Y\left(\theta^t\right)}{B_C\left(\theta^t\right)}$ can still be used to provide an upper or lower bound on the optimal top income tax.

[19] Marginal information rents at any given type are multi-dimensional as they depend on the direction of a possible deviation. Hence there is no natural counter-part to the $A\left(\theta\right)$ component of the ABC-formula, which captured local information rents at a given type, in multidimensional type spaces.

the analysis, and that the core idea of redistributional arbitrage across different dimensions of the commodity space also applies in multi-dimensional type spaces.

6 Conclusion

We developed a new perspective on optimal tax design, based on the idea that optimal allocations trade off not only between efficiency and redistribution, but also between the margins along which redistribution takes place. The optimal tax system then equalizes the marginal benefit of redistribution from higher to lower types for all goods, around any given type θ, a property that we call redistributional arbitrage. As our main result, we derived a particularly simple new formula for optimal tax distortions based on redistributional arbitrage. We show how to infer the respective marginal benefits of redistibution from income and consumption data and key preference elasticities, thus giving empirical content to this new perspective on optimal tax design.

As our main policy implication, our calibration results suggest that there may be significant gains from taxing and redistributing savings at the top of the income distribution. Our model suggests that it may be optimal to tax savings by up to 1.8% per year, while lowering top income taxes substantially relative to existing sufficient statistics calibrations of top income tax rates. These results are consistent with the empirical observation that savings, like income, appear to be far more unequally distributed than consumption, suggesting potential welfare gains from shifting redistribution from consumption towards savings.

The importance of both leisure and consumption for worker welfare is both historically and contemporaneously well documented. This generates trade-offs between different margins of redistributing welfare. Redistributional arbitrage formalizes how these tradeoffs are resolved by optimal tax policies. In practice, many policy makers probably develop an intuitive understanding for redistributional arbitrage, when determining what policies are popular with their voters and matter for the voters' welfare. In fact, the Roman emperors are perhaps the first rulers on record to perform redistributional arbitrage, since they already knew that the most cost-effective way to keep their working population happy was to provide them with a combination of *panem et circenses*, or bread and entertainment![20]

[20]To be fair, the Roman poet Juvenal coined the phrase *panem et circenses* in the early 2nd century to mock the high levels of political corruption, motives that are outside the tradeoffs considered by our benevolent social planner. But what worked for a corrupt Roman politician also works for a benevolent Mirrleesian planner, as long as the working population's welfare depends on being provided the right mix of bread and entertainment.

References

Aguiar, Mark and Erik Hurst (2007). "Measuring trends in leisure: The allocation of time over five decades". In: *The Quarterly Journal of Economics* 122.3, pp. 969–1006.

Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis (2012). "Recent developments in the economics of time use". In: *Annu. Rev. Econ.* 4.1, pp. 373–397.

Atkinson, Anthony Barnes and Joseph E Stiglitz (1976). "The design of tax structure: direct versus indirect taxation". In: *Journal of public Economics* 6.1-2, pp. 55–75.

Auclert, Adrien (2019). "Monetary policy and the redistribution channel". In: *American Economic Review* 109.6, pp. 2333–67.

Badel, Alejandro and Mark Huggett (2017). "The sufficient statistic approach: Predicting the top of the Laffer curve". In: *Journal of Monetary Economics* 87, pp. 1–12.

Browning, Martin and Costas Meghir (1991). "The effects of male and female labor supply on commodity demands". In: *Econometrica: Journal of the Econometric Society*, pp. 925–951.

Chetty, Raj (2006). "A new method of estimating risk aversion". In: *American Economic Review* 96.5, pp. 1821–1834.

— (2009). "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods". In: *Annu. Rev. Econ.* 1.1, pp. 451–488.

— (2012). "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply". In: *Econometrica* 80.3, pp. 969–1018.

Christiansen, Vidar (1984). "Which commodity taxes should supplement the income tax?" In: *Journal of Public Economics* 24.2, pp. 195–220.

Corlett, Wilfred J and Douglas C Hague (1953). "Complementarity and the excess burden of taxation". In: *The Review of Economic Studies* 21.1, pp. 21–30.

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet (2003). "Capital income taxation when inherited wealth is not observable". In: *Journal of Public Economics* 87.11, pp. 2475–2490.

Diamond, Peter (1998). "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates". In: *American Economic Review*, pp. 83–95.

Diamond, Peter and Emmanuel Saez (2011). "The case for a progressive tax: from basic research to policy recommendations". In: *Journal of Economic Perspectives* 25.4, pp. 165–90.

Diamond, Peter and Johannes Spinnewijn (2011). "Capital income taxes with heterogeneous discount rates". In: *American Economic Journal: Economic Policy* 3.4, pp. 52–76.

Diamond, Peter Arthur and James A Mirrlees (1978). "A model of social insurance with variable retirement". In: *Journal of Public Economics* 10.3, pp. 295–336.

Farhi, Emmanuel and Iván Werning (2010). “Progressive estate taxation”. In: *The Quarterly Journal of Economics* 125.2, pp. 635–673.

— (2013). “Insurance and taxation over the life cycle”. In: *Review of Economic Studies* 80.2, pp. 596–635.

Ferey, Antoine, Benjamin Lockwood, and Dmitry Taubinsky (2021). *Sufficient Statistics for Nonlinear Tax Systems with Preference Heterogeneity*. Working Paper. National Bureau of Economic Research.

Gahvari, Firouz and Luca Micheletto (2016). “Capital income taxation and the Atkinson–Stiglitz theorem”. In: *Economics Letters* 147, pp. 86–89.

Gauthier, Stéphane and Fanny Henriet (2018). “Commodity taxes and taste heterogeneity”. In: *European Economic Review* 101, pp. 284–296.

Gerritsen, Aart, Bas Jacobs, Alexandra Victoria Rusu, and Kevin Spiritus (2020). *Optimal taxation of capital income with heterogeneous rates of return*. CESifo Working Paper.

Golosov, Mikhail, Michael Graber, Magne Mogstad, and David Novgorodsky (2021). *How Americans respond to idiosyncratic and exogenous changes in household wealth and unearned income*. Working Paper. National Bureau of Economic Research.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski (2003). “Optimal indirect and capital taxation”. In: *The Review of Economic Studies* 70.3, pp. 569–587.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016). “Redistribution and social insurance”. In: *American Economic Review* 106.2, pp. 359–86.

Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl (2013). “Preference heterogeneity and optimal capital income taxation”. In: *Journal of Public Economics* 97, pp. 160–175.

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin (2014). *A variational approach to the analysis of tax systems*. Working Paper. National Bureau of Economic Research.

Gruber, Jon and Emmanuel Saez (2002). “The elasticity of taxable income: evidence and implications”. In: *Journal of public Economics* 84.1, pp. 1–32.

Hellwig, Christian (2021). *Static and Dynamic Mirrleesian Taxation with Non-separable Preferences: A Unified Approach*. TSE Working Paper.

— (2022a). *Barro meets Mirrlees: Ricardian Equivalence in Dynamic Private Information Economies*. Work in progress.

— (2022b). *Multi-dimensional Screening: a First-Order Approach*. Work in progress.

Jacobs, Bas and Robin Boadway (2014). "Optimal linear commodity taxation under optimal non-linear income taxation". In: *Journal of Public Economics* 117, pp. 201–210.

Jacquet, Laurence and Etienne Lehmann (2021). *How to Tax Different Incomes?* CEPR Discussion Paper No. DP16571.

Keane, Michael P (2011). "Labor supply and taxes: A survey". In: *Journal of Economic Literature* 49.4, pp. 961–1075.

Kocherlakota, Narayana and Luigi Pistaferri (2009). "Asset pricing implications of Pareto optimality with private information". In: *Journal of Political Economy* 117.3, pp. 555–590.

Lewis, Daniel J, Davide Melcangi, and Laura Pilossoph (2019). "Latent heterogeneity in the marginal propensity to consume". In: *FRB of New York Staff Report* 902.

Ligon, Ethan (1998). "Risk sharing and information in village economies". In: *The Review of Economic Studies* 65.4, pp. 847–864.

Mirrlees, James A (1971). "An exploration in the theory of optimum income taxation". In: *The review of economic studies* 38.2, pp. 175–208.

— (1976). "Optimal tax theory: A synthesis". In: *Journal of public Economics* 6.4, pp. 327–358.

Piketty, Thomas and Emmanuel Saez (2013). "A theory of optimal inheritance taxation". In: *Econometrica* 81.5, pp. 1851–1886.

Romer, Christina (2011). *Work-life balance and the economics of workplace flexibility*. DIANE publishing.

Rothschild, Casey and Florian Scheuer (2014). *A theory of income taxation under multidimensional skill heterogeneity*. Tech. rep. National Bureau of Economic Research.

Saez, Emmanuel (2001). "Using elasticities to derive optimal income tax rates". In: *The review of economic studies* 68.1, pp. 205–229.

— (2002). "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes". In: *Journal of Public Economics* 83.2, pp. 217–230.

Saez, Emmanuel and Stefanie Stantcheva (2018). "A simpler theory of optimal capital taxation". In: *Journal of Public Economics* 162, pp. 120–142.

Saez, Emmanuel and Gabriel Zucman (2019a). "How would a progressive wealth tax work? Evidence from the economics literature". In: *Brookings Institution*.

— (2019b). "Progressive wealth taxation". In: *Brookings Papers on Economic Activity* 2019.2, pp. 437–533.

Scheuer, Florian and Joel Slemrod (2021). "Taxing our wealth". In: *Journal of Economic Perspectives* 35.1, pp. 207–30.

Scheuer, Florian and Iván Werning (2016). *Mirrlees meets diamond-mirrlees.* Working Paper. National Bureau of Economic Research.

Schieman, Scott, Philip J Badawy, Melissa A. Milkie, and Alex Bierman (2021). "Work-life conflict during the COVID-19 pandemic". In: *Socius* 7.

Schulz, Karl (2021). *Redistribution of Return Inequality.* CESifo Working Paper.

Spiritus, Kevin, Etienne Lehmann, Sander Renes, and Floris Zoutman (2021). *Optimal Taxation with Multiple Incomes and Types.* Working Paper.

Toda, Alexis Akira and Kieran Walsh (2015). "The double power law in consumption and implications for testing Euler equations". In: *Journal of Political Economy* 123.5, pp. 1177–1200.

Townsend, Robert M (1994). "Risk and insurance in village India". In: *Econometrica: journal of the Econometric Society*, pp. 539–591.

7 Appendix: Proofs and Derivations

Proof of Proposition 1 and Corollary 1:

Define the indirect utility function $W(\theta) = U(C(\theta), Y(\theta); \theta) + \beta V(S(\theta))$. The local IC constraint is given by $W'(\theta) = U_\theta(C(\theta), Y(\theta); \theta)$. The planner's problem is stated as follows:

$$K(v_0) = \min_{\{C(\theta), Y(\theta), S(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \left(C(\theta) - Y(\theta) + R^{-1} S(\theta)\right) f(\theta)\, d\theta, \text{ s.t.}$$

$$\int_{\underline{\theta}}^{\overline{\theta}} W(\theta) f(\theta)\, d\theta \geq v_0$$

$$W(\theta) = U(C(\theta), Y(\theta); \theta) + \beta V(S(\theta))$$

$$W'(\theta) = U_\theta(C(\theta), Y(\theta); \theta).$$

We solve the planner's problem as an optimal control problem using $W(\cdot)$ as the state variable, and $C(\cdot)$, $Y(\cdot)$, and $S(\cdot)$ as controls. Defining λ, $\psi(\theta)$, and $\mu(\theta)$ as the multipliers on respectively the ex ante promise-keeping constraint, the promise-keeping and local IC constraints given θ, the Hamiltonian for this problem is stated as follows:

$$\begin{aligned} \mathcal{H} &= \left\{C(\theta) - Y(\theta) + R^{-1} S(\theta) + \lambda(v_0 - W(\theta))\right\} f(\theta) \\ &\quad + \psi(\theta)(W(\theta) - U(C(\theta), Y(\theta); \theta) - \beta V(S(\theta))) + \mu(\theta) U_\theta(C(\theta), Y(\theta); \theta) \end{aligned}$$

The first-order conditions with respect to the allocations $C(\cdot)$, $Y(\cdot)$, and $S(\cdot)$ yield:

$$\frac{\psi(\theta)}{f(\theta)} = \frac{1}{U_C(\theta)} + \frac{\mu(\theta)}{f(\theta)} \frac{U_{\theta C}(\theta)}{U_C(\theta)} = \frac{1}{-U_Y(\theta)} + \frac{\mu(\theta)}{f(\theta)} \frac{U_{\theta Y}(\theta)}{U_Y(\theta)} = \frac{(\beta R)^{-1}}{V'(S(\theta))}.$$

The first-order conditions for $C(\cdot)$, $Y(\cdot)$, and $S(\cdot)$ define a shadow cost of utility of agents with type θ, $\psi(\theta)/f(\theta)$, which consists of a direct shadow cost $1/U_C(\theta)$, $1/(-U_Y(\theta))$, or $(\beta R)^{-1}/V'(S(\theta))$ of increasing type θ utility through higher consumption, lower earnings or higher savings, and a second term that measures how such a consumption or earnings increase affects $U_\theta(\theta)$ and thereby tightens or relaxes the local incentive compatibility constraint at θ by $\frac{U_{\theta C}(\theta)}{U_C(\theta)}$ or $\frac{U_{\theta Y}(\theta)}{U_Y(\theta)}$. The latter is weighted by the multiplier $\mu(\theta)/f(\theta)$ and added to the former; it is missing from the first-order condition for savings since preferences are separable in savings.

Combining the first two first-order conditions and rearranging terms then yields the following static optimality condition:

$$\frac{1}{U_C(\theta)}\frac{\tau_Y(\theta)}{1-\tau_Y(\theta)} = \frac{1}{-U_Y(\theta)} - \frac{1}{U_C(\theta)} = \left(\frac{U_{\theta C}(\theta)}{U_C(\theta)} - \frac{U_{\theta Y}(\theta)}{U_Y(\theta)}\right)\frac{\mu(\theta)}{f(\theta)} = A(\theta)\frac{\mu(\theta)}{f(\theta)}.$$

The multipliers $\mu(\cdot)$ and λ are derived by solving the linear ODE $\mu'(\theta) = -\frac{\partial \mathcal{H}}{\partial V}$, after substituting out $\psi(\theta)$ using any of the three first-order conditions:

$$\mu'(\theta) = -\frac{\partial \mathcal{H}}{\partial V} = \lambda f(\theta) - \psi(\theta) = \left(\lambda - \frac{1}{U_C(\theta)}\right) f(\theta) - \mu(\theta)\frac{U_{\theta C}(\theta)}{U_C(\theta)},$$

along with the boundary conditions $\mu(\underline{\theta}) = \mu(\overline{\theta}) = 0$. Define $\frac{U_{\theta C}(\theta)}{U_C(\theta)} = \frac{\hat{m}'(\theta)}{\hat{m}(\theta)}$, or $\hat{m}(\theta) = e^{\int_{\underline{\theta}}^{\theta}\frac{U_{\theta C}(\theta')}{U_C(\theta')}d\theta}$. Substituting into the above ODE and integrating out yields

$$\mu(\overline{\theta})\hat{m}(\overline{\theta}) - \mu(\theta)\hat{m}(\theta) = \int_{\theta}^{\overline{\theta}}\left(\lambda - \frac{1}{U_C(\theta')}\right) f(\theta')\hat{m}(\theta')\,d\theta',$$

or

$$\mu(\theta) = \frac{1-\hat{F}(\theta)}{\hat{m}(\theta)}\mathbb{E}(\hat{m}(\theta'))\left\{\hat{\mathbb{E}}\left(\frac{1}{U_C(\theta')}|\theta' \geq \theta\right) - \lambda\right\}.$$

The boundary condition $\mu(\underline{\theta}) = 0$ then gives $\lambda = \hat{\mathbb{E}}(1/U_C(\theta))$. Therefore,

$$\begin{aligned}
\frac{\mu(\theta)}{f(\theta)} &= \hat{H}(\theta)\left\{\hat{\mathbb{E}}\left(\frac{1}{U_C(\theta')}|\theta' \geq \theta\right) - \hat{\mathbb{E}}\left(\frac{1}{U_C(\theta')}\right)\right\} \\
&= \frac{1}{U_C(\theta)}\hat{H}(\theta)\left\{\hat{\mathbb{E}}\left(\frac{U_C(\theta)}{U_C(\theta')}|\theta' \geq \theta\right) - \hat{\mathbb{E}}\left(\frac{U_C(\theta)}{U_C(\theta')}\right)\right\} \\
&= \frac{1}{U_C(\theta)}B_C(\theta).
\end{aligned}$$

Substituting this expression into the static optimality condition then yields the first intra-temporal optimality condition $\frac{\tau_Y(\theta)}{1-\tau_Y(\theta)} = A(\theta)\cdot B_C(\theta)$.

The FOC for earnings yields an analogous ODE,

$$\mu'(\theta) = \left(\lambda - \frac{1}{-U_Y(\theta)}\right) f(\theta) - \mu(\theta)\frac{U_{\theta Y}(\theta)}{U_Y(\theta)}.$$

Apply the same steps as to the first yields

$$\frac{\mu(\theta)}{f(\theta)} = \widetilde{H}(\theta)\left\{\widetilde{\mathbb{E}}\left(\frac{1}{-U_Y(\theta')}|\theta' \geq \theta\right) - \widetilde{\mathbb{E}}\left(\frac{1}{-U_Y(\theta)}\right)\right\} = \frac{1}{-U_Y(\theta)}B_Y(\theta)$$

and the intra-temporal optimality condition $\tau_Y(\theta) = A(\theta)\cdot B_Y(\theta)$ along with $\lambda = \widetilde{\mathbb{E}}(1/-U_Y(\theta))$.

Finally, we solve for the inter-temporal optimality condition. Combining the ODE $\mu'(\theta) = -\frac{\partial \mathcal{H}}{\partial V} = \lambda f(\theta) - \psi(\theta)$ with the FOC for savings yields

$$\mu'(\theta) = \left(\lambda - \frac{(\beta R)^{-1}}{V'(S(\theta))}\right) f(\theta),$$

which can be integrated and solved along the same lines as above to find

$$\frac{\mu(\theta)}{f(\theta)} = H(\theta)(\beta R)^{-1}\left\{\mathbb{E}\left(\frac{1}{V'(S(\theta'))}|\theta' \geq \theta\right) - \mathbb{E}\left(\frac{1}{V'(S(\theta))}\right)\right\} = \frac{(\beta R)^{-1}}{V'(S(\theta))}B_S(\theta).$$

Equating this last expression to $\frac{1}{U_C(\theta)}B_C(\theta)$ then yields the expression for the savings wedge:

$$1 + \tau_S(\theta) \equiv \frac{U_C(\theta)}{\beta R V'(S(\theta))} = \frac{B_S(\theta)}{B_C(\theta)}.$$

Finally, consider that savings are unbounded at the top. The Inada condition on V then implies $\lim_{\theta\to\bar{\theta}} \beta R V'(S(\theta)) = 0$. It follows that $\lim_{\theta\to\bar{\theta}}\left(-U_Y\left(\theta'\right)\right) = \lim_{\theta\to\bar{\theta}} \frac{B_Y(\theta)}{B_S(\theta)}\beta R V'(S(\theta))$. It is straight-forward to check that $\lim_{\theta\to\bar{\theta}} B_S(\theta)/H(\theta) \geq 1$ and $\lim_{\theta\to\bar{\theta}} B_Y(\theta)/H(\theta) \leq 1$, and therefore $\lim_{\theta\to\bar{\theta}}(-U_Y(\theta)) \leq \lim_{\theta\to\bar{\theta}} \beta R V'(S(\theta)) = 0$. Finally, $\lim_{\theta\to\bar{\theta}} U_C(\theta) = \lim_{\theta\to\bar{\theta}} \frac{1}{1-\tau_Y(\theta)}(-U_Y(\theta))$, which equals 0 whenever $\lim_{\theta\to\bar{\theta}} \tau_Y(\theta) < 1$.

Proof of Corollary 2: Suppose that $U_{\theta C}(\theta) \leq 0$. We then have

$$\mu'(\theta) = \left(\lambda - \frac{1}{U_C(\theta)}\right) f(\theta) - \mu(\theta)\frac{U_{\theta C}(\theta)}{U_C(\theta)}$$

and

$$\mu'(\theta) = \left(\lambda - \frac{(\beta R)^{-1}}{V'(S(\theta))}\right) f(\theta).$$

Hence,

$$\frac{(\beta R)^{-1}}{V'(S(\theta))} = \frac{1}{U_C(\theta)} + \frac{\mu(\theta)}{f(\theta)}\frac{U_{\theta C}(\theta)}{U_C(\theta)}.$$

Furthermore, proposition 1 implies that $\mu(\theta) > 0$ for all θ. Recall our assumption that $U_{\theta C}(\theta)$ is either everywhere non-positive, or everywhere non-negative. It then follows immediately that $U_C(\theta) \lesseqgtr \beta R V'(S(\theta))$, or $\tau_S(\theta) \gtreqless 0$ for all θ, if and only if $U_{\theta C}(\theta) \lesseqgtr 0$ $\tau_S(\theta) \gtreqless 0$ for all θ.

Proof of Proposition 2:

Totally differentiating $U_C(\theta)$ yields

$$\begin{aligned}\frac{\frac{d}{d\theta}U_C(\theta)}{U_C(\theta)} &= \frac{U_{C\theta}(\theta)}{U_C(\theta)} + \frac{U_{CC}(\theta)}{U_C(\theta)}C'(\theta) + \frac{U_{CY}(\theta)}{U_C(\theta)}Y'(\theta) \\ &= \frac{U_{C\theta}(\theta)}{U_C(\theta)} - \mathcal{E}_C(\theta)\frac{C'(\theta)}{C(\theta)} + \eta(\theta)\end{aligned}$$

where

$$\eta(\theta) = \frac{Y(\theta)U_{CY}(\theta)}{U_C(\theta)}\frac{Y'(\theta)}{Y(\theta)} - \frac{C(\theta)U_{CY}(\theta)}{-U_Y(\theta)}\frac{C'(\theta)}{C(\theta)} = (1-\kappa(\theta))\frac{Y(\theta)U_{CY}(\theta)}{U_C(\theta)}\frac{Y'(\theta)}{Y(\theta)}.$$

where $\kappa(\theta) = \frac{U_C(\theta)}{-U_Y(\theta)}\frac{C'(\theta)}{Y'(\theta)} = \frac{C'(\theta)}{(1-\tau_Y(\theta))Y'(\theta)}$. It follows that $\hat{m}(\theta) = e^{-\int_\theta^{\overline{\theta}}\frac{U_{\theta C}(\theta')}{U_C(\theta')}d\theta'} = \frac{1}{U_C(\theta)}\hat{M}(\theta)$, where

$$\hat{M}(\theta) = e^{-\int_\theta^{\overline{\theta}}\mathcal{E}_C(\theta')d\ln C(\theta') + \int_\theta^{\overline{\theta}}\eta(\theta')d\theta'} = e^{-\int_\theta^{\overline{\theta}}\mathcal{E}_C(\theta')d\ln C(\theta')}\Psi(\theta),$$

where $\Psi(\theta) = e^{\int_\theta^{\overline{\theta}}\eta(\theta')d\theta'}$. Applying the same steps to $-U_Y(\theta)$ yields

$$\begin{aligned}\frac{\frac{d}{d\theta}(-U_Y(\theta))}{-U_Y(\theta)} &= \frac{U_{Y\theta}(\theta)}{U_Y(\theta)} + \frac{U_{YY}(\theta)}{U_Y(\theta)}Y'(\theta) + \frac{U_{CY}(\theta)}{U_Y(\theta)}C'(\theta) \\ &= \frac{U_{Y\theta}(\theta)}{U_Y(\theta)} + \mathcal{E}_Y(\theta)\frac{Y'(\theta)}{Y(\theta)} + \eta(\theta)\end{aligned}$$

where $\widetilde{m}(\theta) = \frac{1}{-U_Y(\theta)}\widetilde{M}(\theta)$ and $\widetilde{M}(\theta) = e^{\int_\theta^{\overline{\theta}}\mathcal{E}_Y(\theta')d\ln Y(\theta')}\Psi(\theta)$. Using the local IC constraint

$$U_C(\theta)C'(\theta) + U_Y(\theta)Y'(\theta) + \beta V'(S(\theta))S'(\theta) = 0.$$

Proof of Theorem 1:

It follows from the Inada condition and the condition that elasticities converge to finite limits that

$$
\begin{aligned}
\lim_{\theta\to\bar{\theta}} \tau_Y(\theta) &= 1 - \lim_{\theta\to\bar{\theta}} \frac{\mathbb{E}\left(\frac{\widetilde{M}(\theta')}{\widetilde{M}(\theta)}|\theta' \geq \theta\right)}{\mathbb{E}\left(\frac{\hat{M}(\theta')}{\hat{M}(\theta)}|\theta' \geq \theta\right)} = 1 - \lim_{\theta\to\bar{\theta}} \frac{\mathbb{E}\left(e^{-\int_\theta^{\theta'} \mathcal{E}_Y(\theta'') d\ln Y(\theta'')} \frac{\Psi(\theta')}{\Psi(\theta)}|\theta' \geq \theta\right)}{\mathbb{E}\left(e^{\int_\theta^{\theta'} \mathcal{E}_C(\theta'') d\ln C(\theta'')} \frac{\Psi(\theta')}{\Psi(\theta)}|\theta' \geq \theta\right)} \\
&= 1 - \lim_{\theta\to\bar{\theta}} \frac{\mathbb{E}\left(\left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_Y} \frac{\Psi(\theta')}{\Psi(\theta)}|\theta' \geq \theta\right)}{\mathbb{E}\left(\left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}_C} \frac{\Psi(\theta')}{\Psi(\theta)}|\theta' \geq \theta\right)}
\end{aligned}
$$

where $\frac{\Psi(\theta')}{\Psi(\theta)} = e^{\int_\theta^{\theta'} \eta(\theta'') d\theta''}$.

To complete the characterization, we need to determine the limit behavior of $\Psi(\theta)$ or $\eta(\theta)$. If $\mathcal{E}_Y(\theta)$, $\mathcal{E}_C(\theta)$, $\mathcal{E}_{CY}(\theta) = \frac{Y(\theta) U_{CY}(\theta)}{U_C(\theta)}$ and $\mathcal{E}'_{CY}(\theta) = \frac{C U_{CY}}{-U_Y}$ all converge to finite limits, $\frac{\Psi(\theta')}{\Psi(\theta)} \approx \left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_{CY}} \left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}'_{CY}}$ for θ and θ' sufficiently large. From this we obtain

$$
\lim_{\theta\to\bar{\theta}} \tau_Y(\theta) = 1 - \lim_{\theta\to\bar{\theta}} \frac{\mathbb{E}\left(\left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_Y-\mathcal{E}_{CY}} \left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}'_{CY}} |\theta' \geq \theta\right)}{\mathbb{E}\left(\left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}_C+\mathcal{E}'_{CY}} \left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_{CY}} |\theta' \geq \theta\right)}
$$

For the numerator, define $X(\theta) \equiv C(\theta)^{\mathcal{E}'_{CY}} Y(\theta)^{-\mathcal{E}_Y-\mathcal{E}_{CY}}$. We wish to compute $\mathbb{E}\left(\frac{X(\theta')}{X(\theta)}|\theta' \geq \theta\right)$, given that $C(\theta)$, $Y(\theta)$, and $X(\theta)$ are perfectly co-monotonic and C and Y are distributed according to a Pareto distribution with tail coefficients ξ_C and ξ_Y. Let $G(\cdot)$ the cdf of X, and notice that $1 - G(X(\theta)) = 1 - F(\theta)$ and $g(X(\theta)) = f(\theta)/X'(\theta)$. It follows that

$$
\frac{1 - G(X(\theta))}{X(\theta) g(X(\theta))} = \frac{X'(\theta)}{X(\theta)} H(\theta) = \left(\mathcal{E}'_{CY} \frac{C'(\theta)}{C(\theta)} - (\mathcal{E}_Y + \mathcal{E}_{CY}) \frac{Y'(\theta)}{Y(\theta)}\right) H(\theta) = \frac{\mathcal{E}'_{CY}}{\xi_C} - \frac{\mathcal{E}_Y + \mathcal{E}_{CY}}{\xi_Y},
$$

or $X(\theta)$ follows a Pareto distribution with tail coefficient $\left(\frac{\mathcal{E}'_{CY}}{\xi_C} - \frac{\mathcal{E}_Y+\mathcal{E}_{CY}}{\xi_Y}\right)^{-1}$. This implies

$$
\lim_{\theta\to\bar{\theta}} \mathbb{E}\left(\left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_Y-\mathcal{E}_{CY}} \left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}'_{CY}} |\theta' \geq \theta\right) = \frac{1}{1 + \frac{\mathcal{E}_Y+\mathcal{E}_{CY}}{\xi_Y} - \frac{\mathcal{E}'_{CY}}{\xi_C}}
$$

Along the same lines,

$$
\lim_{\theta\to\bar{\theta}} \mathbb{E}\left(\left(\frac{C(\theta')}{C(\theta)}\right)^{\mathcal{E}_C+\mathcal{E}'_{CY}} \left(\frac{Y(\theta')}{Y(\theta)}\right)^{-\mathcal{E}_{CY}} |\theta' \geq \theta\right) = \frac{1}{1 - \left(\frac{\mathcal{E}_C+\mathcal{E}'_{CY}}{\xi_C} - \frac{\mathcal{E}_{CY}}{\xi_Y}\right)},
$$

and therefore

$$\lim_{\theta\to\bar{\theta}}\tau_Y(\theta)=1-\frac{1-\left(\frac{\mathcal{E}_C+\mathcal{E}'_{CY}}{\xi_C}-\frac{\mathcal{E}_{CY}}{\xi_Y}\right)}{1+\frac{\mathcal{E}_Y+\mathcal{E}_{CY}}{\xi_Y}-\frac{\mathcal{E}'_{CY}}{\xi_C}}=\frac{\mathcal{E}_C/\xi_C+\mathcal{E}_Y/\xi_Y}{1+\mathcal{E}_Y/\xi_Y+\mathcal{E}_{CY}/\xi_Y-\mathcal{E}'_{CY}/\xi_C}.$$

At the optimal allocation, $B_C(\theta)$ must be finite, and therefore $\frac{\mathcal{E}_C}{\xi_C}<1+\frac{\mathcal{E}_{CY}}{\xi_Y}-\frac{\mathcal{E}'_{CY}}{\xi_C}$. It then follows automatically that $\lim_{\theta\to\bar{\theta}}\tau_Y(\theta)<1$. To complete the proof we replace $\mathcal{E}'_{CY}/\xi_C=\kappa\cdot\mathcal{E}_{CY}/\xi_Y$.

To prove the second part of Theorem 1, combine $\lim_{\theta\to\bar{\theta}}B_S(\theta)=\frac{1}{1-\mathcal{E}_S/\xi_S}$ for $\mathcal{E}_S/\xi_S<1$ with $\lim_{\theta\to\bar{\theta}}B_C(\theta)=\frac{1}{1-\frac{\mathcal{E}_C}{\xi_C}+(1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}}$ for $1+(1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}>\frac{\mathcal{E}_C}{\xi_C}$ to get $\lim_{\theta\to\bar{\theta}}\tau_S(\theta)=\frac{1-\frac{\mathcal{E}_C}{\xi_C}+(1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}}{1-\frac{\mathcal{E}_S}{\xi_S}}-1=\frac{\frac{\mathcal{E}_S}{\xi_S}-\frac{\mathcal{E}_C}{\xi_C}+(1-\kappa)\frac{\mathcal{E}_{CY}}{\xi_Y}}{1-\frac{\mathcal{E}_S}{\xi_S}}$.

Proofs of Section 4:

From wedges to taxes. A tax schedule $\{T_Y(Y),T_S(S)\}$ implements an allocation $\{C(\theta),Y(\theta),S(\theta)\}$ if and only if the latter maximizes utility subject to the budget constraint, for all types. In particular, this implies that the following two first-order conditions hold along with the budget constraint whenever the allocation is strictly positive:

$$1-T'_Y(Y(\theta))=\frac{-U_Y(\theta)}{U_C(\theta)}\quad\text{and}\quad 1+T'_S(S(\theta))=\frac{\beta RV'(S(\theta))}{U_C(\theta)}.$$

It follows that a tax schedule implements the optimal labor and savings wedges $\{\tau_Y(\theta),\tau_S(\theta)\}$ if and only if $T'_Y(Y(\theta))=\tau_Y(\theta)$ and $T'_S(S(\theta))=\tau_S(\theta)$. Given the profile of wedges, we thus construct the following tax functions:

$$T_Y(Y)=T_Y(Y(\underline{\theta}))+\int_{Y(\underline{\theta})}^{Y}\tau_Y\left(\Theta\left(Y'\right)\right)dY'=T_Y(Y(\underline{\theta}))+\int_{\underline{\theta}}^{\Theta(Y)}\tau_Y(\theta)\,Y'(\theta)\,d\theta$$

$$T_S(S)=T_S(S(\underline{\theta}))+\int_{S(\underline{\theta})}^{S}\tau_S\left(\Theta\left(S'\right)\right)dS'=T_S(S(\underline{\theta}))+\int_{\underline{\theta}}^{\Theta(S)}\tau_S(\theta)\,S'(\theta)\,d\theta$$

where $\Theta(Y)$ and $\Theta(S)$ are the inverses of $Y(\theta)$ and $S(\theta)$. Set $T_S(S(\underline{\theta}))$ and $T_Y(Y(\underline{\theta}))$ so that the budget constraint is satisfied for the lowest type.[21] We write the budget constraint for higher

[21] Notice that the optimal allocation only determines the net present value $R^{-1}T_S(S(\underline{\theta}))+T_Y(Y(\underline{\theta}))$ of the lowest type's tax or transfer, but the timing remains indeterminate. More generally, Ricardian equivalence w.r.t. lump-sum taxes and transfers continues to hold in Mirrleesian economies (Hellwig (2022a), mimeo available upon request).

types as

$$\int_{\underline{\theta}}^{\theta} \left[S'(\theta')\left(1+T_S'(S(\theta'))\right)+R\left\{Y'(\theta')\left(1-T_Y'(Y(\theta'))\right)-C'(\theta')\right\}\right]d\theta' \geq 0$$

Substituting the two first-order conditions for $1+T_S'(S(\theta))$ and $1-T_Y'(Y(\theta))$ and using the local incentive compatibility constraint $U_C(\theta)C'(\theta)+U_Y(\theta)Y'(\theta)+\beta V'(S(\theta))S'(\theta)=0$ then implies that the budget constraint is satisfied for all types. Hence the proposed labor and savings taxes indeed implement the optimal labor and savings wedges, provided that second order conditions are also verified.[22]

Finally, note that the timing of tax collection is also indeterminate. Hence it would be possible for the planner to defer income taxes until the second period (but still tax based on first period income), or levy savings taxes early along with income taxes, i.e., when agents decide to save. An equivalent implementation of the optimal allocation sets a tax payment in period 1 equal to $T(Y(\theta)) \equiv T_Y(Y)+\frac{1}{R}(S(Y)+T_S(S(Y)))$ along with a promised savings payment $S(Y(\theta))=S(\theta)$, while imposing such high levies on additional private savings that no agent saves or borrows on their own. This tax system implements the optimal allocation without any private savings. The optimal tax schedule can then be interpreted as a combination of income taxes $T_Y(Y)$ and social security contributions $\frac{1}{R}(S(Y)+T_S(S(Y)))$, with the savings wedge mapping into the excess or shortfall of social security contributions over savings or pension payments.

Mapping from elasticities $\mathcal{E}_Y, \mathcal{E}_C, \mathcal{E}_{CY}, \mathcal{E}_S$ to income and substitution effects. A perturbation $(\delta\hat{T}_Y, \delta\hat{T}_S)$ of the tax system leads to responses $(\delta\hat{Y}, \delta\hat{C}, \delta\hat{S})$ by the agents that satisfy the perturbed first-order conditions:

$$-\frac{U_Y\left[C+\delta\hat{C}, Y+\delta\hat{Y}, \theta\right]}{U_C\left[C+\delta\hat{C}, Y+\delta\hat{Y}, \theta\right]} = 1-T_Y'(Y)-\delta\left\{\hat{T}_Y'(Y)+T_Y''(Y)\hat{Y}\right\}$$

and

$$\frac{\beta R V'\left(S+\delta\hat{S}\right)}{U_C\left[C+\delta\hat{C}, Y+\delta\hat{Y}, \theta\right]} = 1+T_S'(S)+\delta\left\{\hat{T}_S'(S)+T_S''(S)\hat{S}\right\}$$

with

$$\hat{C}+\left(1+T_S'(S)\right)\hat{S}+\hat{T}_S(S)=\left(1-T_Y'(Y)\right)\hat{Y}-\hat{T}_Y(Y).$$

[22]Intuitively, the second-order conditions hold whenever the two tax functions are convex, or marginal taxes progressive, or U and V are sufficiently concave to make up for non-convexities in the tax functions.

Suppose that the tax schedules are locally linear in the top bracket, so that $T''_Y(Y) = T''_S(S) = 0$. We obtain the responses of earnings, consumption and savings by taking first-order Taylor expansions of the two perturbed FOCs as $\delta \to 0$. Tedious but straightforward algebra leads to:

$$\frac{\hat{Y}}{Y} = -\frac{1}{\mathcal{E}_Y}\frac{\hat{T}'_Y(Y)}{1-T'_Y(Y)} - \frac{\mathcal{E}_C}{\mathcal{E}_Y}\frac{\hat{C}}{C}$$

and

$$\begin{aligned}\frac{\hat{C}}{C} = &-\frac{\frac{\pi_Y}{\pi_S\gamma_S}\mathcal{E}_S}{\frac{\gamma_C}{\pi_S\gamma_S}\mathcal{E}_S + \frac{\gamma_C}{\pi_Y}\mathcal{E}_{CY}}\left[\frac{\hat{T}_Y(Y)}{\left(1-T'_Y(Y)\right)Y} + \frac{\pi_S\gamma_S}{\pi_Y}\frac{\hat{T}_S(S)}{\left(1+T'_S(S)\right)S}\right] \\ &+ \frac{1}{\frac{\gamma_C}{\pi_S\gamma_S}\mathcal{E}_S + \frac{\gamma_C}{\pi_Y}\mathcal{E}_{CY}}\left[\frac{\hat{T}'_Y(Y)}{1-T'_Y(Y)} + \frac{\hat{T}'_S(S)}{1+T'_S(S)}\right] + \frac{\pi_Y}{\gamma_C}\left[1 + \frac{\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y}{\frac{\gamma_C}{\pi_S\gamma_S}\mathcal{E}_S + \frac{\gamma_C}{\pi_Y}\mathcal{E}_{CY}}\right]\frac{\hat{Y}}{Y}.\end{aligned}$$

Solving this linear system of two equations and two unknowns (ignoring for conciseness the perturbation of the capital taxes) leads to the following correspondence between the two sets of elasticities:

$$\zeta_Y^{sub} = \frac{\frac{1/\mathcal{E}_Y}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)} + \frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{\frac{\pi_S\gamma_S}{\pi_Y}/\mathcal{E}_S}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}{1+\frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_Y/\mathcal{E}_S}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}, \quad \zeta_Y^{inc} = \frac{\frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{1}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}{1+\frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_Y/\mathcal{E}_S}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}},$$

$$\zeta_C^{sub} = \frac{\frac{1/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}}{1+\frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_Y/\mathcal{E}_S}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}, \quad \zeta_C^{inc} = \frac{\frac{1/\frac{\gamma_C}{\pi_Y}}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{1}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}{1+\frac{\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}{1+\mathcal{E}_C/(\frac{\gamma_C}{\pi_Y}\mathcal{E}_Y)}\cdot\frac{\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_Y/\mathcal{E}_S}{1+\frac{\pi_S\gamma_S}{\pi_Y}\mathcal{E}_{CY}/\mathcal{E}_S}}.$$

Inverting these expressions leads to those given in the text.

Defining the three cases. Differentiating the inter-temporal budget constraint w.r.t. θ and rearranging terms yields

$$\frac{C'(\theta)/C(\theta)}{Y'(\theta)/Y(\theta)}\gamma_C(\theta) + \frac{S'(\theta)/S(\theta)}{Y'(\theta)/Y(\theta)}\frac{1}{R}\gamma_S(\theta)\pi_S(\theta) = \pi_Y(\theta).$$

In addition, if income, consumption and savings have an upper Pareto tail with tail coefficients ξ_Y, ξ_C, and ξ_S, then $\lim_{\theta\to\bar{\theta}}\frac{C'(\theta)/C(\theta)}{Y'(\theta)/Y(\theta)} = \frac{\xi_Y}{\xi_C}$ and $\lim_{\theta\to\bar{\theta}}\frac{S'(\theta)/S(\theta)}{Y'(\theta)/Y(\theta)} = \frac{\xi_Y}{\xi_S}$. Hence we have that

$$\frac{\xi_Y}{\xi_C}\cdot\lim_{\theta\to\bar{\theta}}\gamma_C + \frac{\xi_Y}{\xi_S}\cdot\frac{1}{R}\lim_{\theta\to\bar{\theta}}\gamma_S = 1 \text{ and } \lim_{\theta\to\bar{\theta}}\gamma_C + \frac{1}{R}\lim_{\theta\to\bar{\theta}}\gamma_S = 1.$$

Straight-forward manipulation of the inter-temporal budget constraint implies that $\min\{\xi_C, \xi_S\} = \xi_Y$.[23] This in turn leaves three possible scenarios: (i) $\xi_S > \xi_C = \xi_Y$, (ii) $\xi_C > \xi_S = \xi_Y$, and (iii) $\xi_C = \xi_S = \xi_Y$. Solving for $\lim_{\theta\to\bar{\theta}} \gamma_C$ yields $\lim_{\theta\to\bar{\theta}} \gamma_C = 1$ in case (i), $\lim_{\theta\to\bar{\theta}} \gamma_C = 0$ in case (ii), while in case (iii), $\lim_{\theta\to\bar{\theta}} \gamma_C$ can take on any value between 0 and 1.

Calibration Details: Case (iii) with $\mathcal{E}_{CY} > 0$:

Note that $\kappa \equiv \gamma_C \frac{\xi_C}{\xi_Y} = \gamma_C$ and, by equation (6), $\zeta_C^{inc} = \frac{1}{\gamma_C} \frac{dC}{dT_Y}$. Since $\mathcal{E}_Y = \frac{\zeta_C^{inc}}{\zeta_Y^{inc}} \mathcal{E}_C$, this implies $\mathcal{E}_Y = \frac{1}{\gamma_C \zeta_Y^{inc}} \frac{dC}{dT_Y} \mathcal{E}_C$. This gives us a first expression for $\mathcal{E}_Y$. We also know that $\frac{1}{\mathcal{E}_Y} = \zeta_Y^{sub} + \frac{\zeta_Y^{inc}}{\zeta_C^{inc}} \zeta_C^{sub}$, which gives us a second expression for $\mathcal{E}_Y$. Manipulating these two expressions gives us $\zeta_C^{sub} = \frac{1-\mathcal{E}_Y \zeta_Y^{sub}}{\mathcal{E}_C}$. Next, the equation $\mathcal{E}_{CY} = \frac{1-\zeta_Y^{inc} - \zeta_Y^{sub} \mathcal{E}_Y}{\zeta_Y^{sub} - \frac{\gamma_C}{\pi_Y} \zeta_C^{sub}}$ delivers a third expression: $\mathcal{E}_Y = \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} + \left(\frac{\gamma_C \zeta_C^{sub}}{\zeta_Y^{sub}} - 1\right) \mathcal{E}_{CY}$. Manipulating the third and second expressions leads to $\mathcal{E}_Y = \left[\frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} + \left(\frac{\gamma_C}{\zeta_Y^{sub}} - \mathcal{E}_C\right) \frac{\mathcal{E}_{CY}}{\mathcal{E}_C}\right] / \left[1 + \frac{\gamma_C \mathcal{E}_{CY}}{\mathcal{E}_C}\right]$. Equating this equation to the first expression for $\mathcal{E}_Y$ finally yields:

$$\frac{\mathcal{E}_{CY}}{\mathcal{E}_C} = \frac{\frac{1}{\gamma_C \zeta_Y^{inc}} \frac{dC}{dT_Y} \mathcal{E}_C - \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}}}{\frac{\gamma_C}{\zeta_Y^{sub}} - \left(1 + \frac{1}{\zeta_Y^{inc}} \frac{dC}{dT_Y}\right) \mathcal{E}_C}.$$

Now, Chetty (2006) estimates $\mathcal{E}_{CY} = \alpha \hat{\mathcal{E}}_C$, with $\alpha \leq 0.15$. Since $\hat{\mathcal{E}}_C = \mathcal{E}_C + \gamma_C \mathcal{E}_X$, this equality can be rewritten as

$$\frac{\mathcal{E}_{CY}}{\mathcal{E}_C} = \frac{\alpha}{1-\gamma_C \alpha}.$$

Equating the right hand sides of the previous two equations yields

$$\frac{\zeta_Y^{inc}}{\zeta_Y^{sub}} \alpha \gamma_C^2 + \left(\frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} - \mathcal{E}_C \alpha\right) \gamma_C - \frac{1}{\zeta_Y^{inc}} \frac{dC}{dT_Y} \mathcal{E}_C = 0$$

The positive root of this equation is given by

$$\gamma_C = \frac{\mathcal{E}_C \alpha - \frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} + \sqrt{\left(\frac{1-\zeta_Y^{inc}}{\zeta_Y^{sub}} - \mathcal{E}_C \alpha\right)^2 + \frac{4\alpha}{\zeta_Y^{sub}} \frac{dC}{dT_Y} \mathcal{E}_C}}{2 \frac{\zeta_Y^{inc}}{\zeta_Y^{sub}} \alpha}$$

For $\alpha = 0.15$ and $\mathcal{E}_C \in \{0.5, 0.75, 1, 1.5\}$, we get $\gamma_C \in \{0.182, 0.277, 0.373, 0.574\}$ and, by the previous derivations, $\mathcal{E}_{CY} = \frac{\alpha}{1-\gamma_C \alpha} \mathcal{E}_C \in \{0.078, 0.116, 0.159, 0.244\}$ if $\xi = 2$. We then obtain $\mathcal{E}_Y = \frac{1}{\zeta_Y^{inc}} \frac{dC}{dT_Y} \frac{\mathcal{E}_C}{\gamma_C} \in \{2.198, 2.166, 2.145, 2.09\}$ and $\mathcal{E}_S = \frac{(1-\gamma_C)\zeta_Y^{inc}}{\zeta_Y^{sub} - \gamma_C \zeta_C^{sub}} \in \{0.866, 0.784, 0.69, 0.49\}$. For

[23] If $\xi_C < \xi_Y$ ($\xi_S < \xi_Y$), then consumption (savings) shares grow arbitrarily large, which violates that γ_C and $\frac{1}{R}\gamma_S$ are both bounded between 0 and 1. If $\min\{\xi_C, \xi_S\} > \xi_Y$, then γ_C and $\frac{1}{R}\gamma_S$ both converge to 0, which violates the inter-temporal budget constraint.

$\alpha = 0.1$, we get $\gamma_C \in \{0.181, 0.273, 0.367, 0.56\}$ and $\mathcal{E}_{CY} \in \{0.051, 0.077, 0.104, 0.159\}$. We then obtain $\mathcal{E}_Y \in \{2.21, 2.198, 2.18, 2.143\}$ and $\mathcal{E}_S \in \{0.86, 0.77, 0.679, 0.485\}$.

Deriving the over-identifying restriction (8). To derive the overidentifying restriction given in the text, differentiate the individual's first-order conditions with respect to θ (assuming constant marginal tax rates at the top) to get

$$\begin{aligned}
\frac{CU_{CY}}{U_Y}\frac{C'(\theta)}{C(\theta)} + \frac{YU_{YY}}{U_Y}\frac{Y'(\theta)}{Y(\theta)} + \frac{U_{Y\theta}}{U_Y} &= \frac{CU_{CC}}{U_C}\frac{C'(\theta)}{C(\theta)} + \frac{YU_{CY}}{U_C}\frac{Y'(\theta)}{Y(\theta)} + \frac{U_{C\theta}}{U_C} \\
\frac{CU_{CC}}{U_C}\frac{C'(\theta)}{C(\theta)} + \frac{YU_{CY}}{U_C}\frac{Y'(\theta)}{Y(\theta)} + \frac{U_{C\theta}}{U_C} &= \frac{SV''(S)}{V'(S)}\frac{S'(\theta)}{S(\theta)}.
\end{aligned}$$

Multiplying both sides by $H(\theta)$, forming their ratio and taking the limit as $\theta \to \infty$ yields equation (8).