Journal of Public Economics 235 (2024) 105139

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Inequality as an externality: Consequences for tax design☆

Morten Nyborg Støstad [a,b,*], Frank Cowell [c]

[a] NHH Norwegian School of Economics, Norway
[b] U.C. Berkeley, United States of America
[c] London School of Economics and Political Science, United Kingdom

ARTICLE INFO

JEL classification:
H21, H23, D62, D63

Keywords:
Inequality
Externalities
Optimal income taxation
Welfare

ABSTRACT

Economic inequality may affect a wide range of societal outcomes, for example crime rates, economic growth, and political polarization. In this paper we discuss how to model such effects in welfarist frameworks. Our main suggestion is to treat economic inequality itself as an externality, which has wide-ranging implications for classical economic theory. We show this through the Mirrlees (1971) optimal non-linear income taxation model, where we focus on a post-tax income inequality externality. Optimal top marginal tax rates are particularly affected by the externality, implying a novel equality dimension to optimal top tax rate design. We propose that inequality's externality properties may have larger optimal top tax rate implications than standard revenue concerns; our model thus provides a theoretical basis for real-world governmental tax choices that seem irrational under standard optimal taxation models. We also show that the total inequality aversion implied by the current U.S. tax system is insufficient to accommodate both social welfare weights that are decreasing in income and a significant concern for inequality's externality effects.

1. Introduction

Suppose that economic inequality has societal consequences, changing for example the crime rate, the amount of social unrest, or the political polarization in society. Individuals are then affected by changes in economic inequality even if their own income or wealth is unchanged. As we all affect economic inequality through market actions, it follows that economic inequality is an externality.

In this paper we explore the implications of this idea in both a general fashion and within the Mirrlees (1971) optimal non-linear income taxation model. We develop the concept of an inequality externality, which in practice implies an inequality metric in the individual's utility function. This is reminiscent of other-regarding preferences, but with the crucial distinction that even narrowly self-interested individuals are affected. While our paper thus has many connections to the other-regarding preference literature, we are closest to the previous works exploring how economic inequality's societal effects could affect welfare optimization. The most notable of these is Thurow (1971), which showed that the First Welfare Theorem fails if the income distribution is a pure public good, but also of note are Alesina and Giuliano (2011), where inequality's effects on society are modeled as affecting consumption and thus utility, and Rueda and Stegmueller (2016), where crime is considered a negative externality of inequality.

Despite these contributions, the link between inequality's societal effects and broader economic thought has remained relatively unclear. Considering inequality as an externality allows us to clarify this connection. We combine the extensive bodies of knowledge on economic inequality and economic externalities to streamline what has traditionally been a complex set of phenomena into one simple, tractable concept that builds on existing economic theory.

To show the potentially large consequences of an inequality externality in economic theory we use the Mirrlees (1971) optimal income taxation model. The Mirrlees model is a pillar of modern public economics (Diamond, 1998; Saez, 2001) which centers on the equity-efficiency trade-off from tax revenue collection. On one hand, any

☆ We thank Stéphane Gauthier, Marc Fleurbaey, Emmanuel Saez, Laurence Jacquet, Daniel Waldenström, Olof Johansson-Stenman, Fredrik Carlsson, Etienne Lehmann, Eddy Zanoutene, and Karine Nyborg for helpful comments and discussions. We have also benefited from suggestions from Marie Young Brun, Max Lobeck, Elif Cansu Akoğuz, Stefanie Stantcheva, Thomas Blanchet, Antoine Bozio, François Fontaine, Bertil Tungodden, Damián Vergara, Thomas Piketty, and seminar participants at the Paris School of Economics, U.C. Berkeley, the University of Oslo, the GT Économie de la Fiscalité, ECINEQ 2019, the 2021 EEA Congress, LAGV 2021, the 2021 IIPF Annual Congress, and the 2021 NTA Annual Conference on Taxation. Morten Nyborg Støstad is thankful to the Paris School of Economics and the Aix-Marseille School of Economics, where a large portion of this work was conducted. Finally, we are deeply grateful to the Journal of Public Economics' Editor Nathaniel Hendren and two anonymous referees for invaluable comments and suggestions.

* Corresponding author at: NHH Norwegian School of Economics, Norway.
E-mail addresses: morten.stostad@nhh.no (M.N. Støstad), f.cowell@lse.ac.uk (F. Cowell).

https://doi.org/10.1016/j.jpubeco.2024.105139
Received 6 August 2022; Received in revised form 26 April 2024; Accepted 8 May 2024
Available online 22 May 2024

small tax increase mechanically raises tax revenues absent behavioral changes, thus allowing for more redistribution. On the other hand, those same behavioral responses generally decrease work effort, thus limiting the possible redistribution. The welfare effects of this trade-off, which can be calculated under various assumptions, center around the resulting income changes. In the standard model, individual welfare only depends on the individual's own income. This is because agents are indifferent to other individuals' incomes by assumption, which in turn implies that individuals are unaffected by the income distribution itself. If inequality has societal consequences, this may not be an appropriate simplification.

We explore the ramifications of the assumption by introducing an income inequality externality into the Mirrlees model. The consequences of introducing externalities into the Mirrlees framework have been extensively analyzed in the context of other-regarding preferences (Oswald, 1983; Aronsson and Johansson-Stenman, 2008; Kanbur and Tuomala, 2013; Aronsson and Johansson-Stenman, 2018) and environmental externalities (Sandmo, 1975; Bovenberg and van der Ploeg, 1994; Cremer et al., 1998). We add to this literature by introducing an externality based on income inequality into the second-best optimal income taxation model. A similar goal is also achieved by the concurrently developed (Aronsson and Johansson-Stenman, 2023), which explores a broader mathematical model in a distinct conceptual setting. We discuss this further below.

In a technical sense, our framework adds a consumption-based externality to the Mirrlees model. The externality has a varying marginal effect that is affected by the full distribution of consumption levels. We make several assumptions to make progress, most notably imposing separability in income, work effort, and the inequality externality.[1] As such, the level of income inequality has no effect on the individual's labor supply. We also use *absolute* inequality metrics to allow inequality to be societally relevant across changing societal income levels,[2] and our main specification assumes no income effects in either labor effort or the marginal cost of inequality. These simplifications allow us to sharply focus on the social planner's new equality incentives, and thus to intuitively understand the main drivers of optimal tax changes when income equality itself becomes a policy goal.

Within this framework we calculate optimal marginal tax rates analytically and numerically in the presence of various types of inequality externalities. While we focus on a post-tax income inequality externality based on the Gini coefficient, we also introduce other types of inequality externalities into the model (pre-tax income, utility) and vary the inequality metric itself. To pin down plausible magnitudes of a real-world income inequality externality we utilize three distinct methods, the primary of which is survey data from Carlsson et al. (2005) and all of which imply similar magnitude ranges. Finally, we perform an inverse-optimum exercise to examine how implied social welfare weights in the U.S. tax system change in the presence of an income inequality externality.

We present two main results within this optimal income taxation (OIT) framework. First, top marginal tax rates are particularly sensitive to the inequality externality. This can be seen through the two consequences of a small tax increase; the mechanical effect and agents' behavioral responses. In the revenue-maximizing case, these two effects oppose each other across the distribution. The same is not true for the impact on income inequality. The mechanical effect always reduces post-tax income inequality by redistributing tax revenue away from those in higher tax brackets (except at the very bottom tax bracket). This is in-line with the classic "equity" effect. Agents' behavioral responses, meanwhile, increase or decrease post-tax income inequality depending on the location of the tax hike. At the bottom, a behavioral shift away from work effort *in*creases income inequality. At the top, a behavioral shift away from work effort *de*creases income inequality. This creates a distributional asymmetry; the optimal tax effects of the mechanical effect and agents' behavioral responses always oppose each other near the bottom of the distribution and harmonize near the top. Theoretically, then, it appears that top marginal tax rates should be particularly sensitive to the magnitude of the inequality externality.

This theoretical intuition is borne through in our numerical simulations. Applying an inequality externality based on the Gini coefficient in post-tax income with a baseline magnitude calibrated with survey data from Carlsson et al. (2005) – which elicits respondents' income-inequality trade-off in a hypothetical setting – results in the optimal top marginal tax rate increasing from 63% to 81%. This estimate implies a certainty associated with the inequality externality magnitude that does not exist in practice, however. As such, we explore the optimal tax consequences of a range of inequality externality magnitudes – which leads to a wide range of potentially optimal top marginal tax rates. Given standard parameter values and reasonable inequality externality magnitudes we find optimal top marginal tax rates ranging from negative (<0%) rates if inequality is a positive externality to extremely high (>90%) rates if inequality is a negative externality. This range is wider than what is supported by standard parameter values in the no-externality case, where optimal top marginal tax rates can range between approximately 50% and 80% (most frequently being estimated at 65%–70%). We thus provide a theoretical basis for previously unsupported policy arguments, such as the post-war top marginal tax rates in the United States and the United Kingdom, both above 90%. Combined with the theoretical intuition above, we propose that inequality's externality properties may have larger optimal top tax rate implications than standard revenue concerns.

Our second main result comes from the inverse-optimum exercise popularized by Bourguignon and Spadaro (2012). This method calculates the implied social welfare weights (SWWs) of real-world tax systems under the assumption that the tax schedule was set optimally. As shown in Lockwood and Weinzierl (2016) and Hendren (2020), SWWs from the U.S. tax schedule are generally decreasing in income in the no-externality case. We introduce an inequality externality into this framework. At the core of this exercise is a substitution effect between the two redistributive motives (decreasing SWWs and a negative inequality externality). If we suppose that the social planner considered inequality as a negative externality when designing the tax schedule this must imply less progressive SWWs *given the same (actual) tax schedule*. This allows us to explore the possible motives behind the total inequality aversion in the U.S. tax schedule. Through this exercise we find our second main result; the 2019 U.S. tax system is not sufficiently inequality averse to accommodate both SWWs that decrease in incomes and a concern about the societal effects of economic inequality equivalent to less than half of our baseline estimate. While the tax system can accommodate either redistributive motive, it cannot accommodate both. If the U.S. social planner considered inequality as an externality to our baseline value, implied SWWs are sharply increasing in income – indicating that the social planner values one dollar at the bottom of the distribution equally to five dollars at the top. We conclude that the current U.S. tax schedule either has increasing SWWs in income, or has an implied concern for inequality's externality effects that is significantly lower than both our empirical estimates and U.S. citizens' concerns as detailed by Lobeck and Støstad (2023).

Broadly speaking, the optimal tax analysis shows that an inequality externality significantly changes welfare modeling conclusions. While this is unsurprising given the existing literature on externalities and other-regarding preferences, the conceptual framework we present indicates a novel reason to be concerned about such fragility. We thus

[1] The assumption of separability in externalities is weakened by among others Pirttilä and Tuomala (1997) and Jacobs and De Mooij (2015). For more on the separability assumption see Gauthier and Laroque (2009).

[2] Absolute inequality metrics are equal to traditional inequality metrics (such as the Gini coefficient) scaled to average income. We discuss this further below.

spend the last section of the paper further developing the concept of the inequality externality.

We first clarify the conditions under which an inequality externality exists by creating simple micro-foundations for various inequality externality channels. In doing so we show that inequality can be an externality even if individuals are narrowly self-interested with no knowledge of the true income distribution. Assuming that incomes causally and monotonically change political opinions along one dimension is enough to micro-found a direct effect of income inequality on political polarization, for example. We then explore what other theoretical implications an inequality externality may have in welfare frameworks, making both general comments and briefly discussing three recent works where conclusions may be particularly affected by an inequality externality – Guvenen et al. (2019), Heathcote et al. (2020), and Jacquet and Lehmann (2021). In general, modeling implications are likely to be particularly large when the policy lever strongly affects economic inequality.

As the core part of our paper relates to optimal taxation, we will now briefly outline how our model differs from the existing OIT literature. Our contribution can be clearly placed relative to three key papers; Oswald (1983), Kanbur and Tuomala (2013), and Aronsson and Johansson-Stenman (2023).

First, both Oswald (1983) and Kanbur and Tuomala (2013) examine the effect on the optimal tax rate when average income is an externality. Our OIT framework is mathematically speaking a specific case of the most general form of these models. The externality structure we focus on, however – based on inequality and not average income – is not explicitly explored by Oswald (1983) or Kanbur and Tuomala (2013). Most of our results, including the distributional asymmetry we discuss above, is a direct consequence of this focus. We also focus on a small-perturbation framework to build intuition, unlike both these papers which use mechanism design frameworks (the modified version of which we also solve).

Aronsson and Johansson-Stenman (2023) is most closely related to our work. While the main focus of our paper is to develop the conceptual idea of inequality as an externality, Aronsson and Johansson-Stenman (2023) focuses on the mathematical implications of various types of other-regarding preferences within the Mirrlees model. Still, one of their formulations is mathematically equivalent to a Gini-based post-tax income inequality externality. This formulation expands on certain mathematical limitations of our work, for example by allowing individuals' labor decisions to depend on the level of income inequality. Their paper thus provides a complement to our approach. Nonetheless, our specific focus allows for various advantages; we explore the effect of varying the relevant inequality metric, numerically estimate the Gini-based inequality externality magnitude through three distinct methods, show the underlying theoretical intuition with the small-perturbation framework, and present an inverse-optimum exercise. We also use absolute inequality metrics, or traditional inequality metrics such as the Gini coefficient multiplied by the average income, to ensure that inequality's externality properties remain relevant across societal income levels when using additive utility functions.[3]

Beyond the differences already outlined we make three technical contributions to the OIT literature. First, we solve the inverse-optimum problem (Bourguignon and Spadaro, 2012) in the presence of a global externality. Global externalities are rarely discussed in this literature – we are only aware of Tsyvinski and Werquin (2017), who discuss the compensation principle in a general equilibrium-based framework and is thus both conceptually and mathematically different from our work. Given the large focus on inequality's effects on society in political rhetoric, we believe this is a particularly interesting exercise in our framework. Second, we explore the optimal taxation effects of pre-tax income inequality and utility inequality externalities in the Mirrlees framework. The introduction of a pre-tax income inequality externality leads to a marginal tax schedule that largely increases in income, resembling real-world tax schedules. And third, we show that a specific family of inequality metrics, which includes the Gini coefficient, is particularly suitable for optimal taxation problems. The formulation changes analytically intractable inequality metrics into a linear combination of consumption externalities with varying marginal effects that depend only on the income-rank of the individual. This simplifies an analytically intractable externality problem into a relatively simple form. As such we can use much of the existing externality framework, including the aforementioned Oswald (1983) and Kanbur and Tuomala (2013), to evaluate what would otherwise be a challenging analytical problem. This is the same family applied to social welfare functions in Simula and Trannoy (2022), developed concurrently with this paper.

Finally, we note three other particularly related works. Kaplow (2010) mentions that the economic distribution could affect variables such as crime, which could imply optimal taxation effects; Aronsson and Johansson-Stenman (2018) explores the first-best Pareto-efficient marginal tax structures under inequality-averse agents; and Kanbur et al. (1994) examines direct poverty concerns in the social welfare function.

The paper is organized as follows. Section 2 briefly summarizes the existing evidence on inequality's societal consequences. Section 3 examines the concept of inequality as an externality and how it differs from other ways in which distributional concerns are modeled in conventional OIT analysis. Section 4 incorporates an inequality externality into a standard OIT model and investigates the impact of the externality on optimal tax rates. Section 5 conducts numerical simulations in the OIT framework. Section 6 discusses the inequality externality concept further, creating micro-foundations and discussing other potential mathematical formulations. Section 7 concludes.

2. Inequality's societal consequences: Evidence

How does economic inequality affect various facets of society and thus individuals' lives? It is difficult to establish causality on the topic for several reasons, the first among them being the lack of exogenous variation in macroeconomic inequality. Other empirical concerns include measurement error, reverse causality (where outcomes also affect inequality), non-linear effects of inequality on outcomes, and more. Still, there is no shortage of academic papers on the subject, some of which we review here.

Small-scale experiments, which avoid most of the above problems, have recently argued that economic inequality between workers or experimental subjects could negatively impact productivity (Breza et al., 2018), trust (Fehr et al., 2020), cooperation (Xu and Marandola, 2022), and other outcomes. External validity is difficult to ascertain, however. In macroeconomic settings, a large collection of works has attempted to measure various cross-country correlations of inequality and societal outcomes. While the related literature is too large to summarize here, examples of relevant reviews can be found in Rufrancos et al. (2013) for crime and Bergh et al. (2016) for individual health. Broadly speaking, such works often find that economic inequality is associated to negative outcomes. We show two examples of such correlations in Fig. 1 for generalized trust and homicides.

It is also notable that both laypeople and experts often express the belief that inequality does change society. In the United States, the large

[3] To explain why this is necessary, suppose $U = x - \eta\bar{\theta}$ where x is income, η is a constant, and $\bar{\theta}$ is a standard income inequality metric. Here the social planner is incentivized to increase incomes to reduce the relative importance of inequality even if actual inequality $\bar{\theta}$ is unchanged. This is due to the different units of income (on the scale of x) and income inequality (between 0 and 1). Our approach solves this problem and instead implies a model where inequality's externality impacts remain equally relevant across different historical time periods (as the $\bar{\theta}$ we use is on the scale of x). A functionally identical but somewhat more mathematically cumbersome approach would be to use standard inequality metrics and endogenously multiply η by the average income.

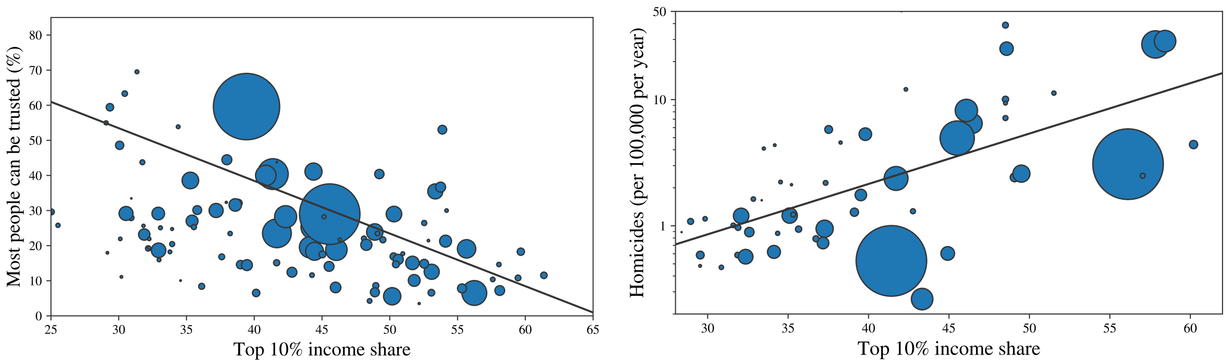

Fig. 1. The correlations of inequality.
Note: Left: The cross-country correlation of generalized trust (World Values Survey) and the top 10% income share (World inequality database). Right: The cross-country correlation of homicides (World Bank) and the top 10% pre-tax national income share (World Inequality Database). Both correlations remain strong when controlling for average income. Data point area is proportional to population. Note the logarithmic scaling of homicide rate. Calculations by the authors.

majority of citizens believe that economic inequality negatively affects a wide range of societal outcomes (Lobeck and Støstad, 2023). Similar concerns have been raised by prominent politicians, philosophers, and economists.[4] Laboratory experiments also indicate that a majority of individuals would forego part of their income to live in more macroeconomically equal societies (Carlsson et al., 2005; Fisman et al., 2021; Bergolo et al., 2022).

From a theoretical angle it is trivial to create realistic microfoundations of various inequality externalities, which we show in Section 6. Other papers have given more attention to specific potential channels; Benabou (1996), Auclert and Rognlie (2018), Mian et al. (2020), and Jones (2022) are just a few examples. We note that these channels do not necessarily have to be societally harmful; Jones (2022) discusses how top incomes could drive innovation, for instance.

In sum, the existing evidence indicates that economic inequality most likely has at least some societal consequences. Still, there remains considerable uncertainty around what these consequences are as well as how they affect net social welfare. In the remainder of the paper we will thus largely remain agnostic on these points, instead showing the policy conclusions that follow from inequality itself having a varying associated net cost or benefit.

3. Inequality and social welfare: An externality approach

What is the optimal approach for incorporating the societal consequences of inequality into a welfare framework? Suppose that economic inequality $\bar{\theta}$ causally affects non-consumption goods individuals care about, the relevant of which we capture in an vector $\overrightarrow{\Psi_i}$. The most natural example of such goods are public goods (such as the quality of democratic institutions), but they might also be individual-specific (such as individual health) – see Section 6.1 for a further discussion on various channels. Suppose further that economic inequality can affect individual consumption x_i (Alesina and Giuliano, 2011),[5] and that individuals may have other-regarding preferences over economic inequality (Cooper and Kagel, 2016).[6] The individual's utility can thus be written as,

$$U_i(x_i(\bar{\theta}), \bar{\theta}, \overline{\overrightarrow{\Psi_i}}(\bar{\theta}), \ldots). \tag{1}$$

Detailed information on each component in the specification (1) is unlikely to be available, and such complexity would also be unrealistically cumbersome for most models. We propose a simplification, noting that the separate contributions are less important than the overall impact of inequality in the utility function. The specification (1) could be written more compactly as the simplified form:

$$\tilde{U}_i(x_i, \bar{\theta}, \ldots) \tag{2}$$

where $\tilde{U}_i$ is the modified utility function and the term $\bar{\theta}$ represents the total impact of the inequality externality on the individual.[7]

The simplication from (1)→(2), which we discuss further in Appendix A, does not rely on the existence of any of the three components we show in (1). The externality exists as soon as one of the three components enters the utility function and is deemed policy-relevant. For instance, individuals could be wholly self-serving and still have a utility function that is strongly dependent on economic inequality if economic inequality affects some pertinent public good. Given the many philosophical problems with introducing ORP and thus emotions into the welfarist framework – as discussed by Harsanyi (1977) and Goodin (1986), among others – this scenario may often be appropriate, and we focus on it for the remainder of the article.

Before we continue, however, it is worth noting that the simplified form (2) is mathematically equivalent to an ORP term in the utility function. It follows that many of the results from the ORP literature can be applied to our framework. This immediately hints at the potential practical significance of the inequality externality, as ORP modifications often have large impacts on standard model conclusions (e.g. Oswald, 1983; Kanbur and Tuomala, 2013).

The concept also needs a well-defined inequality metric $\bar{\theta}$. We return to this later in the paper, but we note that the main type of inequality we will focus on is *income* inequality. We note that optimal capital and wealth taxation models could be heavily affected by a wealth inequality externality. For simplicity we avoid other concerns that, while nonetheless important, complicate a first approach. These issues include questions related to perceived inequality, inequality in different regions, (non-)meritocratic inequality, and so on.

We will now make a short detour to discuss how the inequality externality fits into the general utilitarian framework. In such models

[4] For example Plato (est. 360): "In a state which is desirous of being saved from the greatest of all plagues [...] here should exist among the citizens neither extreme poverty, nor, again, excess of wealth, for both are productive of both these evils", translated in Plato (2016). More recent examples include Greenspan (2014): "You can see the deteriorating impact of [inequality] on our current political system", or Obama (2011): "This kind of inequality – a level that we haven't seen since the Great Depression – hurts us all".

[5] Alesina and Giuliano (2011) discusses how income inequality could affect the income of individuals through three channels; externalities in education, crime and property rights, and incentive effects. One could also imagine that individual income is affected through some of the other channels we discuss in this work (political capture, innovation, social unrest, and so on).

[6] The overbar indicates a society-wide variable.

[7] We have assumed that the parts of $\overrightarrow{\Psi_i}$ that are not determined by $\bar{\theta}$ are irrelevant for policy analysis.

the social planner maximizes a social welfare function consisting of some weighted sum of every individual's utility. In addition to the inequality externality, there are thus two other channels through which inequality-related concerns can enter into the formulation of social welfare comparisons. These are (i) the cumulative effect of diminishing marginal utilities of income (DMUI), and (ii) social welfare weights (SWWs). We posit that the inequality externality is mathematically and intuitively distinct from these other two channels; except for special cases which we discuss further in Appendix B, an inequality externality cannot be mathematically captured by the other formulations. The intuition is simple: as with any other externality, an inequality externality introduces a gap between socially and individually optimal decisions. The sub-optimality of individual decisions cannot be approximated by suitable SWWs, as discounting utility is dissimilar from discounting income, and also cannot be approximated by modifications to an individualistic utility function as such modifications would have to depend on other agents' incomes.

To illustrate this we show an example where an inequality externality introduces a useful modeling feature. Suppose we have a simple welfare-maximizing model. There is an unbounded increase in the top agent's income. In the standard case, such an increase implies an unambiguously non-negative welfare change. If inequality is an externality, there is instead a salient trade-off between the individual income gain and potential externality effects. We further discuss standard modeling methods, including income-based SWWs (e.g. Saez and Stantcheva, 2016), in Appendix B.

We will now explore the effect of introducing three types of inequality externalities – pre-tax income, post-tax income, and utility – into the Mirrlees (1971) framework.

4. Optimal income taxation: Theory

We consider the second-best solution for a non-linear optimal income taxation schedule with a unit size continuum of individuals in the presence of an inequality externality. Agents are indexed by i. Individual i incurs disutility from pre-tax income (labor) $z_i \geq 0$, derives utility from consumption $x_i(z_i) \geq 0$, and is affected by the society-wide post-tax income inequality $\bar{\theta} > 0$. Pre-tax income z is distributed with a strictly positive density $h(z)$ and cumulative density $H(z)$ across the population. The social planner sets an income tax $T(z)$ dependent on pre-tax incomes such that $x_i = z_i - T(z_i)$. Gathered revenue is issued as a flat dividend to all individuals, where we note that any change from a flat dividend would be equivalent to changing the marginal tax schedule. Utility is quasi-linear in consumption and in the inequality externality so that only consumption is subject to income effects.[8] Individuals' earnings decisions do not depend on the level of income inequality.

The inequality externality implies a utility loss $\eta_i \bar{\theta}$ for the inequality metric $\bar{\theta}$ and the individual-specific magnitude $\eta_i = MRS_{x_i \bar{\theta}} = -\frac{dU_i/d\bar{\theta}}{dU_i/dx_i}$. This η_i measures how much consumption the individual would give up for or pay for one unit decrease in the relevant inequality metric. η_i can vary across individuals; if η_i increases in wage-earning ability, for example, then higher-ability individuals have a higher willingness to pay for equality. We assume that η_i does not directly depend on x_i, an assumption we explore in Appendix C.I.

Individual i chooses z_i to maximize;

$$U_i = \mathcal{U}_i\left(z_i - T(z_i) - v_i(z_i; X_i) - \eta_i \bar{\theta}\right),$$

where the function v_i is increasing and strictly convex in z_i, X_i represents individual characteristics (e.g. productivity), and $\mathcal{U}_i$ is increasing and concave. The social planner has social welfare weights g_i indicating the individual income-based benefit of one more unit of consumption to individual i. For the remainder of the paper we will assume that $\mathcal{U}_i$ is taken into account by the social planner's g_i.

To explore the robustness of our main findings, Appendix D solves for the more general utility function $\tilde{U}_i = \mathcal{U}\left(u(x_i) - V(l_i) - \Gamma_i(\bar{\theta})\right)$ in a mechanism design framework.[9] The main intuition we describe below focuses on how inequality itself changes under a marginal tax rate increase – abstracting from utility for as long as possible – and is thus robust across specifications.

The inequality metric $\bar{\theta}$ concerns post-tax income (consumption) inequality. Variants for pre-tax income inequality and utility inequality are explored in Section 5.6. All analysis will be based on *absolute* inequality metrics. Absolute inequality metrics are equal to standard inequality metrics, such as the Gini coefficient, multiplied by the average income. We use such metrics to allow inequality's externality impacts to scale with income levels in additive utility functions. To explain, suppose $U = x - \eta\theta$ where θ is the standard Gini between 0 and 1. In this case the unit of x matters for the trade-off between income and income inequality (and thus optimal taxation). The net effect of this scale dependence is that the social planner has an inequality-based incentive to increase average incomes to make inequality's externality effects relatively less impactful (even if inequality as measured by the Gini remains unchanged). Our approach avoids this scale dependence problem. We discuss this further in Appendix E.

Inequality metrics are often analytically complicated. To simplify the problem we thus use a particular family of inequality metrics found in Cowell (2000). For post-tax absolute income inequality this family has the form,

$$\bar{\theta}(z, H) = \int_{\underline{x}}^{\bar{x}} \kappa(z) x(z) dH(z), \tag{3}$$

where $\kappa(z)$ is the weight of the agent's consumption in the inequality metric.[10] This weight is crucially only dependent on the *rank* of the individual in the distribution. We have used the rank-invariance between pre-tax income z and post-tax income x to specify the weight in terms of z. As x is endogeneous to the tax system and thus difficult to deal with, this key mathematical trick simplifies the problem.[11] We propose that these rank-dependent inequality metrics, also used by Simula and Trannoy (2022), could represent an important simplification in similar problems.

The inequality weight $\kappa(z)$ is non-decreasing, continuous, positive near the top of the income distribution, and negative near the bottom. For example, the (absolute) Gini coefficient in post-tax income has a weight $\kappa_G(z) = 2H(z) - 1$. In the numerical simulations we will also explore other post-tax income inequality metrics based on other types of rank-specific weights $\kappa(z)$ where $\int_{\underline{z}}^{\bar{z}} \kappa(z) dH(z) = 0$, such as those in the Lorenz (Aaberge, 2000) or S-Gini families (Donaldson and Weymark, 1980).

It is worth mentioning that the real-world externality-relevant inequality metric is likely to be a function of several different inequality metrics. To show an example of this, suppose that inequality's effect on crime is dependent on relative poverty and that inequality's effect on political capture is dependent on the proliferation of top incomes. Both relative poverty $\bar{\theta}_p$ and top income proliferation $\bar{\theta}_t$ are distributional metrics, which we represent in our framework by the distributional weights κ_p and κ_t for their respective inequality

[8] We discuss how adding income effects would affect the solution in Appendix C.I.

[9] We have introduced an individual-specific externality function Γ_i, an increasing and concave function u, and an increasing and strictly convex function V. The functions $\mathcal{U}$, u, and V are identical across all individuals.

[10] The non-absolute version of this family presented in Cowell (2000) is divided by the average income.

[11] A similar trick is also crucial in the more general mechanism design approach in Appendix D, where we use the rank-invariance of x_i and wage-earning ability n_i.

measurements $\bar{\theta}_p$ and $\bar{\theta}_t$. Take then an example with separability and homogeneity in these externality effects, such as in the simple example of $U = x - \eta_p\bar{\theta}_p - \eta_t\bar{\theta}_t$ where η_p and η_t indicate externality magnitudes. The total externality effect is $-\eta_p\bar{\theta}_p - \eta_t\bar{\theta}_t = -\left(\eta_p + \eta_t\right)\int_{\underline{z}}^{\bar{z}}\left(\frac{\eta_p}{\eta_p+\eta_t}\kappa_p(z) + \frac{\eta_t}{\eta_p+\eta_t}\kappa_t(z)\right)x(z)dH(z)$. The modified inequality metric is thus $\bar{\theta}_{true} = \int_{\underline{z}}^{\bar{z}}\left(\frac{\eta_p}{\eta_p+\eta_t}\kappa_p(z) + \frac{\eta_t}{\eta_p+\eta_t}\kappa_t(z)\right)x(z)dH(z)$, a weighted sum of the two inequality metrics, with an externality magnitude of $\eta_{true} = \eta_p + \eta_t$. As such, the inequality metrics and externality magnitudes we use could be seen as a combination of potentially several externality-determining inequality metrics.

As we have also introduced potentially heterogeneous inequality magnitudes, we note that our framework allows for a combination of both heterogeneous inequality metrics and heterogeneous inequality magnitudes. The net inequality externality is then $\eta_{net}\bar{\theta}_{net} = \int_j g_j \sum_t \left(\eta_{jt}\bar{\theta}_{jt}\right)dj$, where t indicates the type of externality and j indicates the individual. This allows for individual-specific inequality metrics and externality magnitudes for a flexible number of inequality externalities.

The social planner's chosen tax schedule $T(z)$ is such that no given small perturbation ϵ which changes the tax schedule as $T(z) + \epsilon\Delta T(z)$ leads to welfare improvements. We denote the resulting change in the inequality metric from the small tax increase by $\Delta\bar{\theta}$. The local optimal tax criterion is thus defined as the tax schedule $T(z)$ for which any small budget neutral tax reform in direction $\Delta T(z)$ has $\int_i g_i\left[\Delta T(z_i) + \eta_i\Delta\bar{\theta}\right]di = 0$, where g_i is the SWW of individual i. Although these first-order conditions are only *necessary* criteria for the tax system to be optimal, we assume here that they are also *sufficient*; in every numerical simulation we check that the second-order conditions also hold.

4.1. Optimal marginal tax schedules

Following Saez (2001), we consider an infinitesimally small marginal tax rate increase $\partial\tau(z)$ for individuals in a small band of income between z and $z + dz$ that leaves marginal tax rates unchanged at all other income levels.[12] There are five welfare-pertinent effects of such a change. Three of these are well-known from the previous literature. These are (i) the mechanical effect of higher tax revenue, dM (ii) the behavioral responses of agents reducing their work effort, dB, (iii) and the welfare-relevant income losses of those who are taxed more, dW. There are also two new two equality consequences; (iv) the inequality impact of the mechanical effect, dI_M, and (v) the inequality impact of the behavioral responses, dI_B. At the optimum, the sum of the welfare effect of these five changes must equal to zero:

$$dM + dB + dW + dI_M + dI_B = 0 \tag{4}$$

dM and dB will be discussed as *revenue effects*. The behavioral response dB represents a tax revenue loss, while the mechanical effect dM represents a tax revenue gain. In our set-up the revenue gain from those above z is $[1 - H(z)]dz\partial\tau$. The revenue loss from those in the band is denoted by $-dz\partial\tau \cdot \epsilon(z)zh(z)\tau(z)/(1-\tau(z))$, where $\epsilon(z)$ is the elasticity of earnings $\epsilon(z)$ with respect to $1-\tau(z)$ (see Appendix C for derivation). The two terms together represent a revenue collection trade-off and together form the basis for the calculation of the revenue-maximizing tax rate. In non-Rawlsian SWFs there is also a pertinent welfare loss from the agents above the tax bracket who have their individual incomes reduced, dW. This effect dampens, but cannot cancel, the revenue-based benefit of the mechanical effect due to the assumption of SWWs that are non-increasing in income, and equals $-dz\partial\tau\int_{\{j:z_j\geq z\}} g_j dj$.

The mechanical effect and behavioral responses also impact post-tax income inequality directly. In the following we will assume a negative inequality externality for simplicity.[13] The mechanical (in)equality effect is denoted as dI_M and the (in)equality effect of the behavioral responses is denoted by dI_B. Before calculating the welfare effect, it is convenient to first calculate the effects of these channels on post-tax income inequality; we denote these effects as $d\theta_M$ and $d\theta_B$ respectively. The flat dividend does not affect absolute inequality metrics, and so we can focus on where income is reduced.[14]

We begin with the mechanical (in)equality effect, or $d\theta_M$. The effect of the mechanical income reductions on post-tax income inequality is the same as the classical mechanical revenue effect weighted by the importance of the individuals above z in the inequality metric. In other words, each dollar of additional revenue from an agent at income $z' > z$ from the mechanical effect corresponds to an inequality reduction of $\kappa(z')$. The mechanical effect thus changes income inequality by $d\bar{\theta}_M = -\bar{\kappa}(z)[1-H(z)]dz\partial\tau$, where we have defined the average inequality metric $\kappa(z')$ above z as $\bar{\kappa}(z)[1-H(z)] = \int_{\{j:z'>z\}}\kappa(z')h(z')dj$. The absolute Gini, for example, has $\bar{\kappa}(z) = H(z)$. As $\bar{\kappa}(z) \geq 0$, this effect always reduces income inequality regardless of the tax bracket in question except at the very bottom.[15]

The behavioral responses indicate a reduction in the work effort and thus the income of agents at z. This also affects post-tax income inequality. The (in)equality effect depends on the weight of these individuals in the inequality metric $\kappa(z)$, how much they change their work effort represented by $\epsilon(z)$, and the change in their post-tax income. We show in Appendix C that this is equal to $d\bar{\theta}_B = -\kappa(z)\cdot dz\partial\tau\cdot\epsilon(z)zh(z)$. As $\kappa(z)$ changes signs across the distribution, so does $d\bar{\theta}_B$. At the bottom, behavioral responses increase income inequality. At the top, behavioral responses decrease income inequality. Notably, this means that behavioral responses have a welfare-*positive* dimension at the top under a negative externality. The changing sign of $d\bar{\theta}_B$ across the distribution contrasts with the always negative dB, and is a key difference between the traditional revenue effects and the new equality impacts. We also note that the reliance on the elasticity $\epsilon(z)$ reverses the standard intuition from the revenue channel, where a high elasticity leads to a low tax rate (as the state should keep tax rates low to collect what little revenue they can). In our case, the state might instead prefer to place high tax rates (or subsidies) at the ends of the distribution to increase or decrease inequality as they see fit. We summarize the discussion of the revenue and inequality effects in Table A2.

The total change $d\bar{\theta} = d\bar{\theta}_B + d\bar{\theta}_M$ in the inequality metric leads to a utility loss of $\eta_i d\bar{\theta}$ for individual i. As a result, the total welfare change is $dI = dI_B + dI_M = -\int_j g_j\eta_j(d\bar{\theta}_B + d\bar{\theta}_M)\cdot dj = -(d\bar{\theta}_B + d\bar{\theta}_M)\cdot\int_j\eta_j g_j dj$. This illustrates that the heterogeneous inequality externalities can be weighted by SWWs g_j to become functionally equivalent to a homogeneous inequality externality $\eta = \int_j \eta_j g_j dj$, where η represents the total suffering from a unit change in the inequality metric as calculated by the government.

We can now consider the externality-induced sign change to optimal marginal tax rates as compared to the standard case. At the bottom, where dI_M and dI_B are in opposition (regardless of whether the externality is positive or negative), the welfare effect of a tax increase through the externality dimension is ambiguous. The change to the optimal marginal tax rate due to the externality is thus also ambiguous. At the top, where the signs of dI_M and dI_B harmonize – both are positive (negative externality) or negative (positive externality) – the

12 We show the full calculation with a post-tax income inequality externality in Appendix C.

13 The same intuition with the opposite welfare direction holds for a positive externality.

14 If we were to use *non-absolute* inequality metrics (where flat income increases change the relevant statistic), the intuition would be overall similar with additional terms to correct for these changes in average income.

15 Formally this is due to $\kappa(z)$ being non-decreasing and the assumption that $\int_{\underline{z}}^{\bar{z}}\kappa(z)dH(z) = 0$.

change to the optimal tax rates is unambiguous. Compared to the standard case, the optimal top rates are higher with a negative post-tax income inequality externality and lower with a positive post-tax income inequality externality.

Inserting the values from the preceding discussion into (4) allows us solve for $\frac{\tau(z)}{1-\tau(z)}$ and find that

$$\frac{\tau(z)}{1-\tau(z)} = \eta\kappa(z) + \frac{\eta\overline{\kappa}(z)}{\alpha(z)\epsilon(z)} + \frac{\left(1-\bar{G}(z)\right)}{\alpha(z)\epsilon(z)}. \tag{5}$$

We use the local Pareto parameter $\alpha(z) = \frac{zh(z)}{1-H(z)}$ and the average SWW above z denoted by $\bar{G}(z)$.[16] The last term corresponds to the traditional (Saez, 2001) result under our assumptions. The two new terms are Pigouvian corrections due to the inequality externality and correspond to dI_B and dI_M respectively.

The mechanical effect: $\frac{\eta\overline{\kappa}(z)}{\alpha(z)\epsilon(z)}$. This term in (5) comes from dI_M in (4) and represents the mechanical effect on the agents located above income z. Agents above z face an additional lump-sum tax and see their post-tax incomes decrease. The tax revenue is redistributed uniformly to every agent, so the post-tax income of every agent below z increases. Post-tax income inequality, as a result, decreases. This provides the government an additional incentive to increase tax rates in addition to the standard revenue considerations; assuming a negative (positive) inequality externality, this term unambiguously increases (decreases) the marginal rate in every tax bracket except at the very top and at the very bottom.

How much this increases optimal marginal tax rates at any z depends on the externality magnitude η and the relative average weight of the agents above the tax bracket z in the inequality metric, represented by $\bar{\kappa}(z)$. If these individuals' incomes contribute heavily to the inequality metric on average, their income losses also heavily reduce inequality and thus improve welfare.[17]

The behavioral response: $\eta\kappa(z)$. This term in (5) comes from dI_B in (4) and represents the behavioral responses of the individuals who are located at income z.[18] Agents at income z work less due to the tax increase as they substitute into leisure. Tax revenue is reduced no matter the location of the tax increase. The direction of the equality impact, on the other hand, depends on the location of the tax bracket. The new term incentivizes individuals who make socially suboptimal labor choices to substitute into leisure, keeping their utility relatively high. This does not imply that the social planner wants to punish certain individuals; while the social marginal welfare of *income* can be negative, the social marginal welfare of *utility* is never negative, all else equal (upholding the Pareto principle). The term directly depends on (i) the inequality externality magnitude η, as a larger externality leads to a larger welfare gain from reducing inequality, and (ii) the relative weight $\kappa(z)$ of the agents at z in the inequality metric; how these individuals' incomes contribute to the inequality metric determines how their income losses influence inequality and thus social welfare.

This term cannot be approximated by non-negative income-based SWWs g_i or any utility-based SWWs, see Appendix D. It invalidates three classic results from the literature based on Mirrlees (1971) noted by Sadka (1976) and Seade (1977) – (i) that the optimal marginal tax rate at the top of a bounded distribution should be zero, as it is instead $\tau(z) = \frac{\eta\kappa(z)}{1+\eta\kappa(z)}$ – reducing the income of the top-earner has become a social cost or benefit in itself, (ii) that the optimal marginal tax rate at the bottom should be zero, which is no longer true for similar reasons, and (iii) that the optimal marginal tax rate is bounded between zero and one, as negative optimal rates are possible. These are not new findings in a mathematical sense, as the same is shown for relative income concerns by Oswald (1983).[19] Still, the modifications to the classic OIT results are intuitively appealing given our conceptual framework, and as such we mention them here.

The externality thus introduces two new terms to the optimal tax formula. In general, the new key model variables are the size of the inequality externality (represented by η) and the choice of the relevant inequality metric (represented by κ).

5. Optimal income taxation: Numerical simulations

In this section we use numerical calculations to find optimal marginal tax rates in the presence of a post-tax income inequality externality.

5.1. Numerical specification

We use the mechanism design solution from Appendix D for the numerical specifications, where individuals are on a continuum of wage-earning abilities n with associated density distribution function $f(n)$ and cumulative distribution function $F(n)$. The associated weight in the post-tax income inequality metric is $\kappa(n)$. We assume quasi-linear utility in consumption, a constant labor elasticity E_L, and a separable linear homogeneous inequality externality. This implies the utility function

$$U(x,l,\bar{\theta}) = x - \frac{l^{(1+\frac{1}{E_L})}}{\left(1+\frac{1}{E_L}\right)} - \eta\bar{\theta}, \tag{6}$$

where l is individual labor supply. We will logarithmically scale the SWF to introduce a classical inequality-aversion motive; the Utilitarian case is equal to $W = \int_i log(U_i)di$, for example. The resulting optimal marginal tax rates $t(n)$ at each productivity level n are,

$$\frac{t(n)}{1-t(n)} = \eta\kappa(n) + \eta\left(1+\frac{1}{E_L}\right)\frac{\bar{\kappa}(n)}{\alpha(n)} + \left(1+\frac{1}{E_L}\right)\frac{1}{f(n)n} \times \int_n^\infty \left[1 - \frac{W_{U(p)}}{\lambda}\right] dF(p),$$

where λ is the marginal value of public funds, $\alpha(n)$ is the local Pareto parameter, and $W_{U(p)}$ is the derivative of the SWF with respect to utility (capturing the aforementioned standard inequality aversion).

The underlying wage-earning ability distribution n is found through inverting the observed income distribution using the solution to the individual's maximization problem, following Saez (2001) and others. We use the US Distributional National Accounts micro-files to measure the 2019 U.S. labor income distribution.[20] The NBER TAXSIM model was used to find marginal tax rates on labor income for any given tax unit in the DINA files.[21] The main focus of the numerical simulations

[16] $\alpha(z)$ is a distributional measure which becomes constant in a Pareto distribution. $\bar{G}(z)$ is defined as $\bar{G}(z)(1-H(z)) = \int_{\{j:z_j \geq z\}} g_j dj \big/ \int_j g_j dj$. In the Rawlsian min–max framework, $\bar{G}(z) = 0$. See Saez (2001) for further discussion on these variables.

[17] The standard model parameter values $\epsilon(z)$ and $\alpha(z)$ also appear in this term. It would be misleading to consider these two parameters as "part of" the mechanical effect, however. If the tax rate was equivalently written as a function of $\tau(z)$, then $\alpha(z)\epsilon(z)$ appears in the term for the behavioral responses:

$$\tau(z) = \frac{1+\eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) - \bar{G}(z)}{1+\eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) + \alpha(z)\epsilon(z) - \bar{G}(z)}.$$

This is because the two parameters signify the relative strength of the mechanical and behavioral channels.

[18] We note that $\epsilon(z)$ and part of $\alpha(z)$ originate from this substitution effect despite not entering into the term in (5).

[19] Generally speaking these three results are fragile and change with many small modifications to the model – see Stiglitz (1982) and Saez (2001) for examples.

[20] Described in Piketty et al. (2018), accessed at https://gabriel-zucman.eu/usdina/ on March 22nd 2023.

[21] Described in Feenberg and Coutts (1993), accessed at https://taxsim.nber.org/ on April 20th 2023. More details in Appendix F.I.

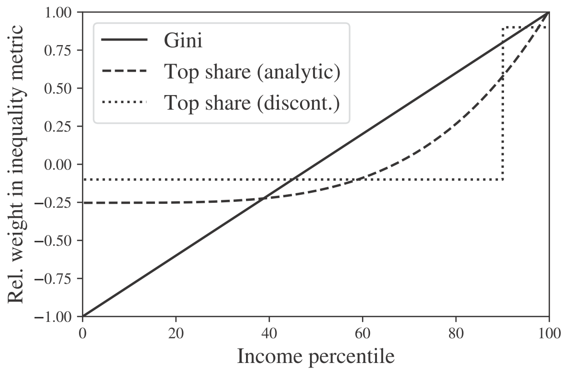

Fig. 2. Weights of inequality metrics.
Notes: Fig. 2 shows the relative weights of individuals' income in the inequality metrics we primarily use (the Gini and the analytic top share metric are used in Fig. 5 and A4, respectively). This corresponds to $\kappa(z)$ in the general expression $\bar{\theta} = \int_{\underline{z}}^{\bar{z}} \kappa(z)x(z)dH(z)$. More inequality metrics are explored in Appendix F.III.

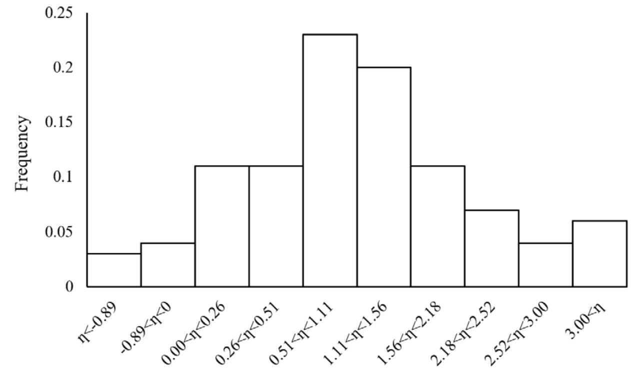

Fig. 3. Estimated η_G.
Notes: Fig. 3 shows the estimated magnitudes of the inequality externality magnitude η_G from the survey experiment in Carlsson et al. (2005). In the following numerical simulations we restrict η_G between −0.5 and 2.0 (and equivalent values for other inequality metrics).

will be on how the inequality externality changes the results from the no-externality case; we thus largely follow the existing literature for the remaining model specification. For more details on the simulation procedure see Appendix F.I.

There are two further choices that are crucial for the simulations that are specific to the inequality externality. These are the choice of the relevant inequality metric $\bar{\theta}$ (e.g. the Gini coefficient in post-tax income) and the magnitude and direction η of the inequality externality. We detail these choices below.

5.1.1. Inequality metric

The inequality metrics we use follow the general form in (3), using wage-earning ability n instead of pre-tax earnings z.[22] In the main specification (Section 5.2) we use the Gini, which has the following form:

$$\kappa_G(n) = 2H(n) - 1. \tag{7}$$

We also show results for a generalized Gini with weights of the following form (see Section 5.3),

$$\kappa_T(n) = (q+1)H(n)^q - 1, \tag{8}$$

which was designed to analytically approximate top income shares (which have a discrete jump and are thus analytically intractable). The Gini corresponds to $q = 1$ in this specification, while larger q approximates top income share inequality metrics by increasingly weighting incomes at the top of the distribution while equalizing the weights of other agents' incomes. The weights of the Gini and the generalized Gini with $q = 4$ are plotted in Fig. 2. We also show the weights used in the top 10% income share for comparison, which is discontinuous and thus not usable in an analytical setting. Other inequality metrics are examined in Appendix F.III.

5.1.2. Inequality externality magnitude

Given the inequality metric we need to choose values for the inequality externality magnitude. The values of η depend on which inequality metric is chosen to be relevant for the externality, and we denote the values calculated for the Gini coefficient as η_G. As there are unavoidable empirical challenges in calibrating such a number,[23] we do not aim to strongly argue for any one value. We instead use a range of realistic η_G to illustrate the potential tax policy consequences of various income inequality externalities. We present three different methods to understand the magnitudes of these η_G.

Correlation-based estimates. To make a reasonable first-pass at an order of magnitude of η_G one could take the cross-country correlation between income inequality and externality dimensions – naïvely taking the correlation after controlling for observables as causal – and use willingness-to-pay estimates for each externality dimension to find the dimension's contribution to the total η_G. We do this for intentional homicides as an illustrative example. We use data from the World Bank for homicides, the World Inequality Database for the Gini coefficient, and Cohen et al. (2004) for the societal willingness to pay for fewer homicides.[24] The correlation between income inequality and intentional homicide is strongly positive, and through this very simple and likely biased approach we find $\eta_{G,homicides} \approx 0.07$.

This only represents a single externality channel, and the full η_G estimate would be found as $\eta_G = \sum_k \eta_{G,k}$. Extending this method to find all $\eta_{G,k}$, however, requires internationally comparable outcome data. This is not a trivial requirement, and precludes the use of more detailed crime data.[25] Other internationally comparable outcomes usually lack willingness-to-pay estimates, making further use of this approach complicated even under the stringent assumptions we impose.

Experimental estimate. To find a range of η_G that takes into account *all* externality dimensions we present estimates based on data from Carlsson et al. (2005). The work uses a survey design to estimate macroeconomic inequality aversion in Swedish university students.[26] The survey, which asks respondents to decide what income-inequality tradeoff their hypothetical grandchildren would prefer, allows us to find individual preferences for η determined to an interval.[27]

[22] These are equivalent as $\kappa(z) = \kappa(n)$ by assumption.

[23] Beyond specific empirical challenges relating to the existence and quality of the available data, it is very challenging – perhaps impossible – to find true exogenous variation in macroeconomic inequality.

[24] Cohen et al. (2004) estimates the total social cost of a homicide as $9.7 million, or $12.8 million corrected for inflation to 2018.

[25] Harrendorf (2018) notes the following: "Crime levels are not a valid measure of crime in different countries, with the possible exception of completed intentional homicide. Total crime rates depend mainly on the internationally differing quality of police work".

[26] Bergolo et al. (2022) finds comparable numbers for Uruguayan university students.

[27] Using a survey experiment instead of a direct externality estimate means that we are relying on potentially biased beliefs to proxy for inequality's externality effects. There is also selection bias in the survey respondents and, because the only degree of freedom is being used to estimate the extent of inequality aversion, it is not possible to know how well our assumed utility function matches the respondents' perceived utility functions. All these reasons contribute to why we are using a *range* of η_G.

Table 1
The magnitude of inequality externalities η_G.

	$\eta_G = -0.5$	$\eta_G = 0.0$	$\eta_G = 0.5$	$\eta_G = 1.0$	$\eta_G = 2.0$	$\eta_G = 3.0$
US income multiple	0.94	1.00	1.06	1.13	1.25	1.38

Notes: Which multiple of their current income would an average-income agent need to move from Denmark-like to U.S.-like inequality? Above are these equivalent incomes for various levels of the inequality externality η_G from the utility function in (6).

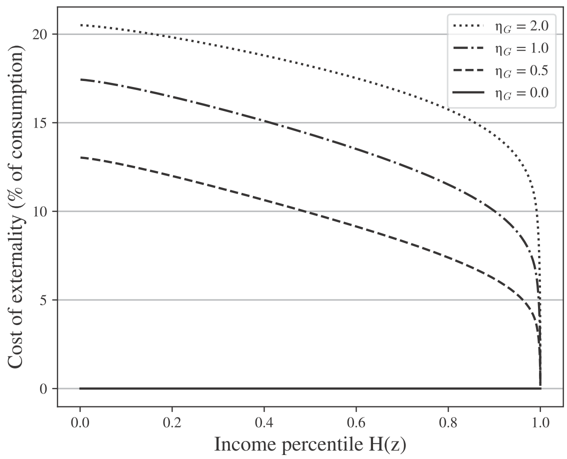

Fig. 4. Cost of the negative Utilitarian inequality externalities in Fig. 5.

The distribution is presented in Fig. 3. The median respondent in the survey has approximately $\eta_G = 1.00$. A majority of respondents have $0.26 < \eta_G < 2.18$.[28] A negative η_G – indicating a preference for inequality, or that inequality is a positive externality – is only observed in 7% of respondents. The equivalent externality magnitude values for top income shares, η_T, are calculated from the same experiment. As a general rule of thumb, $\eta_G \approx 2\eta_T$ when externality magnitudes are equal.

Hypothetical exercises. As these numbers are rather abstract, we present an alternative way of understanding the magnitudes through equivalent incomes. Answering the following question pins down either η: *What multiple of their current income should an average agent require to move from Denmark-like to United States-like inequality?*[29]

Answering the question creates equivalent incomes for the two differing inequality levels, which allows us to pin down an inequality externality magnitude.[30] These equivalent incomes for Denmark and the United States, and their corresponding η_G, are shown in Table 1. As an example, if we have an inequality externality of $\eta_G = 1.0$, the average individual in a society with Denmark's inequality level would require 13% more income to be indifferent if inequality increased to the U.S. level.

Based on these techniques we use the range $-0.5 \leq \eta_G \leq 2.0$ for the Gini-based externality and $-0.15 \leq \eta_G \leq 1.0$ for the top share-based externality in the main numerical simulations.

Evaluating these externality values in the simulations also gives us a way to measure the significance of the inequality externality. In Fig. 4 we show the cost of the inequality externality as the percentage of consumption each income percentile would be willing to give up to remove the externality for each η_G used in the main specification. Although the distribution of the externality impact is not particularly meaningful in our case – the net welfare effect is the policy-determining variable – we believe the illustration gives the reader an idea of the magnitudes we introduce. There are two further caveats to this figure. First, these values are endogenous, as the social planner has already reduced inequality due to the externality. Second, inequality levels are generally very low in optimal income taxation simulations even without an inequality externality; applying the same η_G to real-world inequality levels would mean much higher externality costs.

5.2. Main results: The Gini externalities

Our main specifications, using the Gini as the post-tax income inequality metric, are presented in Fig. 5. To remain general we show both positive and negative inequality externalities. The introduction of even a small post-tax income inequality externality substantially changes the optimal tax structure. The effect is larger towards the top of the income distribution.

We begin by discussing optimal top marginal tax rates, which are equivalent in the Rawlsian and Utilitarian set-ups.[31] With no inequality externality, the optimal top marginal tax rate is 63%. For $\eta_G = 1.00$, the value closest to the empirical externality estimate taken from Carlsson et al. (2005), the optimal top marginal tax rate is 81%. For a somewhat larger negative inequality externality, $\eta_G = 2.0$, the optimal top tax rate increases to 88%. This is higher than the optimal tax rate under the no-externality Rawlsian case, which illustrates that a Rawlsian SWF, in itself, does not imply a maximum dislike of inequality. Meanwhile, a small *positive* inequality externality ($\eta_G = -0.5$) decreases the optimal top marginal tax rate to only 26%. The inequality externality magnitude thus appears to have a large impact on the optimal top tax rate; we will discuss this further in Section 5.4.

The trends across the rest of the distribution differ in the Utilitarian and Rawlsian frameworks. In the Utilitarian framework, a negative (positive) inequality externality shifts the optimal marginal tax rates up (down) across the entire distribution. This is due to the empirical strength of the mechanical effect, which increases (decreases) optimal rates across the entire distribution for a negative (positive) externality. This effect dominates that of the behavioral responses, which increases or decreases optimal rates differentially at the top and bottom, under our parameter choices.[32] The net effects are larger near the top, which is particularly noticeable around the 95th percentile. The larger effects near the top of the distribution is due to the equality effects of the

[28] Due to the design of the experiment, any one individual's inequality aversion is only pinned down to a range.

[29] Assuming the same leisure, that the mean income difference between the two countries is negligible, and that relative position is irrelevant. According to the 2017 World Economic Outlook database GDP per capita is \$61,803 in Denmark, and \$59,707 in the United States. Calculations are based on Gini coefficients of 0.410 for the United States and 0.285 for Denmark.

[30] These numbers are significantly dependent on the income specified (average income in the above case) under a homogeneous inequality externality. They can also be interpreted more generally, however. Under Utilitarianism, quasi-linearity in consumption, and heterogeneous linear and separable inequality externalities, the same percentage can be thought of as the total society-level income increase that would be required for indifference when $\eta = \frac{1}{n}\sum_i \eta_i$. These assumptions correspond to a social welfare function such that $W = \sum_i \left(x_i - \eta_i \bar{\theta}\right)$, as in Sen (1976).

[31] For any given externality, Utilitarian and Rawlsian results converge at the very top (as in the classical literature without externalities). This is due to the assumptions of separability and a homogeneous inequality externality.

[32] This result is not universal, and the effect of the externality at the bottom is usually smaller than in this case due to the counteracting behavioral response. Indeed, the Utilitarian case has among the least top-heavy distributional optimal policy effects of any of our simulations. It is notable that the effects are largest at the top even in this case. Using certain skill distributions, such as the full Pareto distribution in Appendix F.II, a negative externality *decreases* optimal marginal tax rates at the bottom. We also find this result with any pre-tax income inequality externality (see Section 5.6).

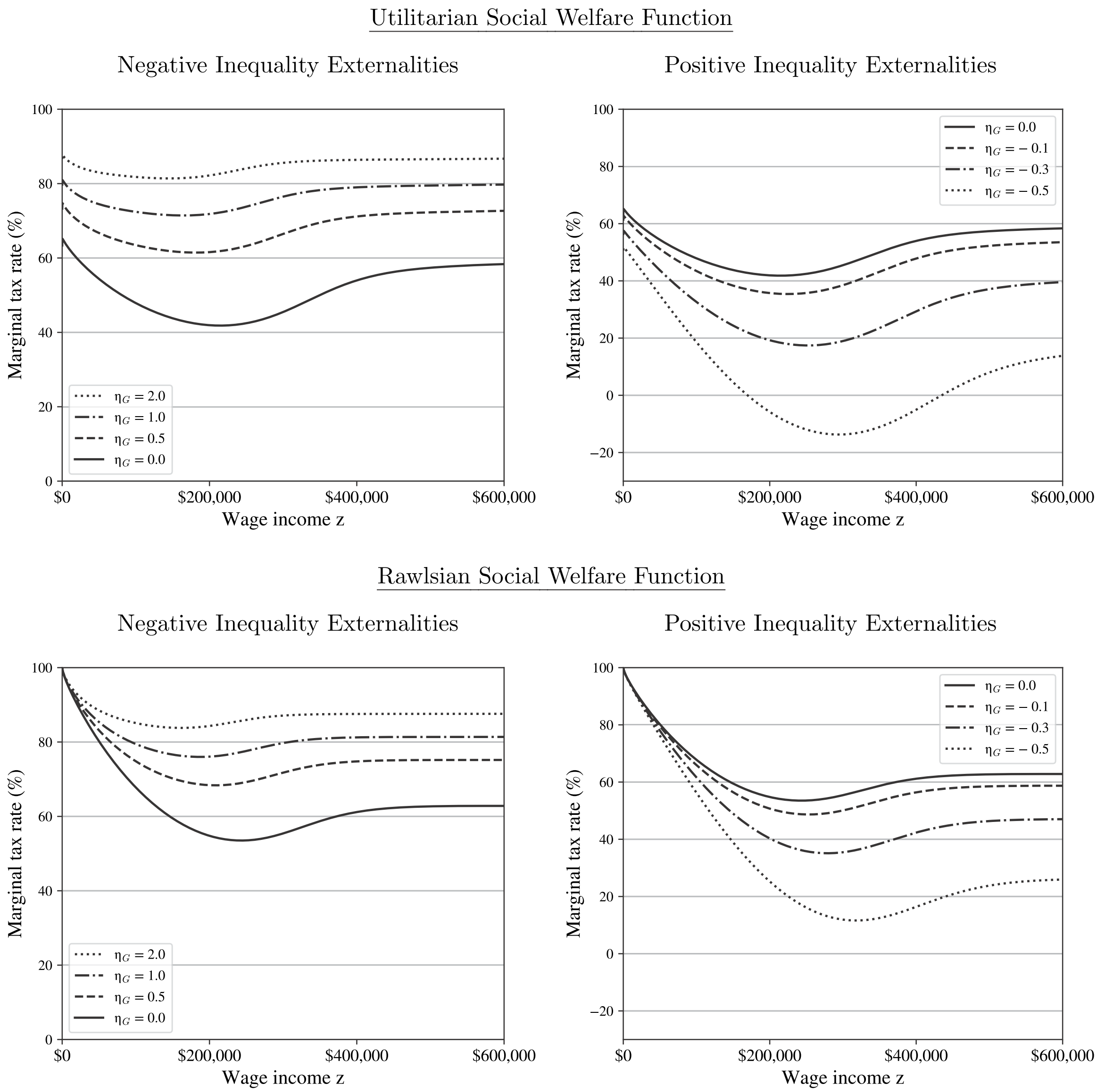

Fig. 5. Optimal marginal income tax schedules with Gini inequality externalities.
Notes: Optimal marginal tax rates for various Gini-based inequality externalities with magnitudes η_G, where inequality is either a negative externality (left) or a positive externality (right). The social planner is Utilitarian (above) and Rawlsian (below). The Utilitarian and Rawlsian cases converge when moving towards the top for a given externality value. Empirical estimates indicate $\eta_G = 1.0$. The solid line, $\eta = 0$, is the standard no-externality case. Further explanation of η is in Table 1. Note the different scales of the vertical axes between the negative and positive externalities.

mechanical and behavioral channels working in the same direction in this region, as discussed in Section 4.1. We also note that all simulations have lower optimal tax rates around the 90th–95th percentiles due to the well-known decrease of the local Pareto parameter around these values, which leads to the classical U-shape found in the literature (Diamond, 1998). In the positive externality case we observe negative optimal tax rates in this region. We return to this shortly.

The Rawlsian externalities we introduce have negligible impacts near the bottom of the distribution, where marginal tax rates are very high in the no-externality case. This is driven by a very high mechanical revenue benefit of taxation near the bottom (which is also found in the classical literature) drowning out any effect of the externality.[33] The effects of the inequality externality are mostly located above the 90th percentile for both negative and positive externalities. Under a positive externality, top marginal tax rates approach zero around the 97th percentile.

The extent of the classical U-shape varies across simulations. It is most striking in the positive externality and no-externality simulations, and is difficult to notice in the negative externality simulations. As the U-shape has been widely discussed as having potential implications for practical tax design it is relevant to ask why this occurs.

[33] The high optimal rates at the bottom of all the Rawlsian simulations are due to the large positive mechanical revenue effects of increasing bottom marginal tax rates. When one only cares about the very bottom agent, as in the Rawlsian case, redistributing away from any other agent is a net positive absent changed labor choices. Since we do not consider income effects, these labor choices do not occur for anyone above the tax bracket in question. The mechanical revenue effect is thus very large at the bottom and leads to very high marginal tax rates in this region. The introduced equality effects are not large enough to change this substantially. In contrast, the Utilitarian simulations take into account the income losses from agents above the tax bracket, which discounts the mechanical benefits of tax increases near the bottom. Very high bottom marginal tax rates are thus less appealing, and the effects of the inequality externality are more visible.

The U-shape emerges from the empirically estimated wage-earning (or income) distribution, as the local Pareto parameter α is high around these wage (or income) percentiles. In short this implies a relative over-density of individuals *in* these tax brackets compared to those *above* these tax bracket, which in turn implies that the relative strength of the behavioral channel is high in this bracket (as compared to the relatively low strength of the mechanical effect). In other words, optimal tax policy in these brackets is increasingly set by the welfare consequences of agents' behavioral responses. This decreases the no-externality optimal tax rates in the region. How does this change when one introduces an inequality externality? In the negative externality case, there is a welfare-positive dimension to the behavioral responses (namely decreased inequality). It follows that an increased importance of the behavioral responses does not necessarily imply a U-shape and lower optimal tax rates – as we can see in the simulations.[34] In the positive externality case, meanwhile, the shift towards a concern for behavioral responses is still highly relevant, as the behavioral responses remain entirely welfare-negative (through decreased revenue and decreased inequality). To summarize, the classical U-shape from the optimal taxation literature may depend on the absence of a negative income inequality externality.

The exact optimal tax structure depends heavily on the model specification, and the numerical simulations should be interpreted with caution.

5.3. Robustness: Top income share externalities

The choice of the inequality metric naturally influences our results. While the Gini coefficient is analytically appealing, it is often considered to over-weight middle-income inequalities. To address this concern we present a robustness check of our main findings where we use the general top income share metric family $\kappa(z) = (q+1)H(z)^q - 1$, $q \in \mathbb{N}$ as the relevant inequality measurement, with increasingly larger q. After $q = 1$, which defines the Gini coefficient, this inequality metric becomes increasingly top-focused and approximates top income share metrics.

We show a set of such inequality metrics and the effect of using them in the optimal tax calculation in Fig. 6. The externality in the optimal tax calculation is kept constant at the median result from Carlsson et al. (2005).[35]

When we move away from the Gini towards a top income share, the effects of the externality are increasingly concentrated towards the top of the distribution. This should not be surprising given the increasing weight of top incomes in the inequality metric, although the *magnitude* of the effect is large. The inequality metric defined by $q = 15$ coupled with the median inequality externality from Carlsson et al. (2005) leads to optimal top marginal tax rates above 95%.

It is also noticeable that the effects near the bottom are reduced. This is not as obvious, as lower inequality metric weights near the bottom have opposite optimal tax effects through the behavioral channel (through which lower κ_{bottom} leads to higher τ) and the mechanical effect (through which lower κ_{bottom} generally leads to lower τ through a higher $\bar{\kappa}_{bottom}$).[36] In the numerical simulations, the mechanical effect is more powerful, indicating that the average marginal externality above is more impactful than marginal externality of the tax bracket itself. Due to this, tax rates for the majority of Americans would be closer to the no-externality case under inequality metrics that focus more on top income shares.

Overall, using top income shares further concentrates the effect of the externality towards the top of the tax schedule. With other inequality metrics, such as those in the S-Gini family, results are overall similar. This is further discussed in Appendix F.III. In sum, the Gini is a conservative choice which dampens effects at the top in return for larger changes across the rest of the distribution. We will now discuss implications for top tax rates specifically.

5.4. Equality concerns: Top tax rates

As we have discussed in the preceding sections, the new equality concerns have a particularly large effect on the optimal top tax rate. The optimal tax rate near the top in the small-perturbation framework with a Gini post-tax income inequality externality is,

$$\tau(z) = \frac{1 + \eta + \eta\alpha(z)\epsilon(z)}{1 + \eta + \eta\alpha(z)\epsilon(z) + \alpha(z)\epsilon(z)}, \tag{9}$$

which is strongly dependent on the inequality externality magnitude η. It is useful to discuss why this occurs.

Revenue considerations, which in this context implies the direct individual effects from the redistribution of income, have few distributional biases. In a Rawlsian set-up, for instance, one tax dollar raised remains one tax dollar raised, regardless of which tax-payer pays it (if not taken from the very bottom).[37] Equality concerns are naturally different: *where* the income is taken from is of key importance. And, as we have seen, the tax policy effects of these equality concerns generally increase as one approaches the top of the distribution.

This implies that the optimal tax rate can be above the revenue-maximizing rate (the so-called "Laffer rate"). The revenue-maximizing rate is occasionally used as an upper bound for sensible tax rates. For example, Piketty et al. (2014) states that they "focused on the revenue-maximizing top tax rate, which provides an upper bound on top tax rates". This position would need to be modified in a model with societal effects of inequality. We discuss this further in Appendix G.

5.4.1. Large variation in top rates: A maximum income, or the Rawlsian conservative?

Some of the variation in international tax brackets, particularly at the top, could be due to policy setters' differing considerations of the inequality externality. Two Rawlsian governments might agree on the elasticity of earnings and revenue-maximizing tax rates and still strongly disagree on optimal top tax rates – *if* they disagree on how inequality changes society. Indeed, varying the value of the inequality-sensitivity parameter η_G has a larger effect on optimal top tax rates than varying the standard parameter values $1/\alpha$ or E_L, which we show in Tables A3 and A4. By changing η_G within reasonable bounds, the same Rawlsian social planner can find optimal top tax rates from close to zero to over 90%. Under stronger positive externalities the same social planner can even find negative optimal top rates.

In other words, a wide range of top tax rates can be optimal depending on the magnitude of the inequality externality. This contrasts with standard OIT models, where top marginal income tax rates usually converge to around 60−70% regardless of the underlying SWF. Although these numbers depend heavily on parameter specifications, heterodox assumptions are required for optimal rates below 50% or above 80%.[38]

[34] Optimal marginal tax rates can even increase in the region under different specifications. In Section I.I this occurs under a negative pre-tax income inequality externality.

[35] The actual values of η change, as estimating η from the Carlsson et al. (2005) data requires an assumption about which inequality metric to use. Changing this inequality metric also changes the calculated η.

[36] In the case of the behavioral channel, the bottom-earner imposes less of an externality and the negative Pigouvian term is thus smaller. In the case of the mechanical effect, redistributing from everyone above is less impactful for inequality-reduction if everyone in the lower half is weighted relatively equally.

[37] In general the welfare changes from a tax and its associated revenue across the distribution is dependent on the SWF. However, the net distributional biases are mechanically constrained due to the non-negativity of the SWWs.

[38] Piketty et al. (2014) finds revenue-maximizing rates varying from 57% to 83% with differing elasticity compositions, for instance.

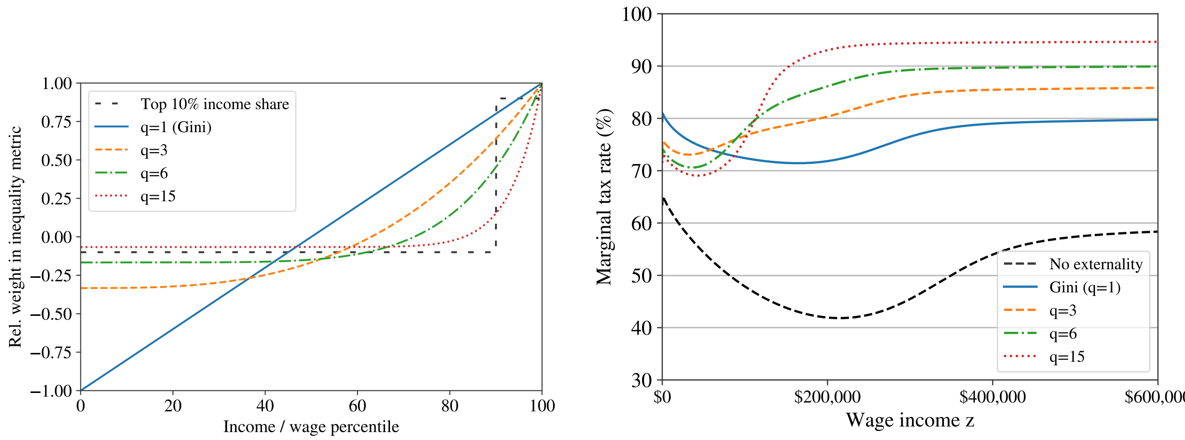

Fig. 6. Varying the inequality metric with a fixed externality magnitude.
Notes: Left: The income weights over the distribution of various inequality metrics in the family where $\kappa(z) = (q+1)F(n)^q - 1$, $q \in \mathbb{N}$. The top 10% income share is also plotted. Larger q indicates that top incomes are increasingly weighted. Right: Optimal marginal tax rates for these inequality metrics, keeping the magnitude of the inequality externality constant for all q at the upper bound of the median value from the empirical inequality aversion estimates in Carlsson et al. (2005). The no-externality case is shown as a reference in dotted black. The wage-earning ability distribution is the empirical income distribution, and the SWF is Utilitarian.

Our model thus rationalizes a wider array of tax schedules. We use two real-world examples to illustrate the power of such a finding.

First, the idea of extremely high top tax rates (a "maximum income"). If one believes in a large negative inequality externality, the negative effect of top income earners on the rest of society is sufficient to argue for top tax rates above 90%. These are similar to tax rates from the post-war period in the United Kingdom, Germany, and the United States. The disincentive for high earners at this stage begins to approach a maximum income.

Second, the idea of a Rawlsian government with low tax rates on the highest income-earners. If one believes in even a small positive inequality externality, here represented by $\eta_G = -0.5$, marginal rates at the top quickly fall below 50% and begin approaching zero. We call this the Rawlsian conservative; the argument that a low top tax rate will lead to the highest possible utility for the worst-off agent.

Both of these intuitive arguments are occasionally discussed in the public sphere. In the standard OIT literature, however, they are unfounded. One strength of our model is that such arguments can be logically substantiated, and disagreements can be traced back to the variable η. Individual opinions on η could be related to (or even determinants of) political leanings and policy preferences.

5.5. U.S. social welfare weights with an inequality externality

As shown in Bourguignon and Spadaro (2012), it is possible to calculate the implied SWWs of the observed tax schedule given the relatively large assumption that the social planner is welfare-maximizing under the constraints of the optimal income tax problem.[39] This method is applied to the U.S. in Lockwood and Weinzierl (2016) and Hendren (2020), both of which generally find decreasing SWWs with income. Hendren (2020), which has more granular data, also notes an increase in SWWs towards the very top of the distribution.

These methods implicitly assume that no inequality externality is taken into account by the social planner when setting the tax schedule. However, U.S. citizens generally believe that inequality has negative consequences (Lobeck and Støstad, 2023). Such beliefs have also been voiced by prominent U.S. politicians.[40] It is thus natural to think that some concern for inequality itself could be included in the income tax schedule design. If so, under the same assumptions from Section 4, we show in Appendix H that the implied SWW $g(z)$ is,

$$g(z) = -\frac{1}{h(z)}\frac{d}{dz}\left[(1-H(z))(1+Y(z)) - \frac{\tau(z)}{(1-\tau(z))} zh(z)\epsilon(z)\right], \tag{10}$$

which differs from the standard case by $Y(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z)$.[41] Intuitively, the implied inequality aversion in a given tax system can come from either the SWF $g(z)$ or externality motivations $Y(z)$, and there is a substitution effect between these two motivations. If externality motivations to avoid inequality were greater when designing a given tax schedule, the same tax schedule will imply that the SWWs in the design process were less progressive.

In Fig. 7 we show $g(z)$ of the 2019 U.S. tax system under standard specifications, assuming the social planner has taken into account various negative post-tax Gini income inequality externalities. The model specification is further discussed in Appendix H.

The standard case of no inequality externality ($\eta_G = 0$) has generally decreasing welfare weights with income with an upward bend towards the top of the distribution, similar to Hendren (2020). Introducing a negative inequality externality ($\eta_G > 0$) changes implied SWWs quickly, however. Implied SWWs are relatively flat for $\eta_G = 0.25$, indicating that all the inequality aversion in the tax system is accounted for by such an inequality externality. The full linear trend is flat at roughly $\eta \approx 0.21$, which can be interpreted as the social planner's estimated inequality externality if all redistributive policy is in actuality driven by an inequality externality.[42] The implied SWWs are increasing for

[39] This is an unlikely assumption, as discussed in Lockwood and Weinzierl (2016). Nonetheless, it is useful to see how current tax systems can be rationalized in the framework of optimal taxation.

[40] For example Obama (2011): "This kind of inequality – a level that we have not seen since the Great Depression – hurts us all".

[41] A few technical points: We use the income density directly, as in Lockwood and Weinzierl (2016), instead of the "virtual" earnings density, as employed in Hendren (2020) and the rest of this work. Due to this the elasticity we use is technically defined to include the circularity between the "virtual" earnings density and the observed income density (Jacquet et al., 2013). This is unlikely to significantly change results due to the absence of pronounced bunching in the actual U.S. income distribution (Saez, 2010). We assume no income effects and no extensive margin behavioral responses for simplicity. A more detailed approach for the no-externality case can be found in Jacobs et al. (2017), which also notes that these factors are empirically small.

[42] No Gini-based income inequality externality yields an exactly flat social welfare weight schedule, so there are different ways to calculate this number. We can also calculate the η_G which corresponds to $\bar{G}(z_{median}) = g(z_{median})$, for example, which indicates the η_G where the average social welfare weight above the median earner is the same as the social welfare weight of the median earner. The corresponding externality magnitude is $\eta_G \approx 0.28$.

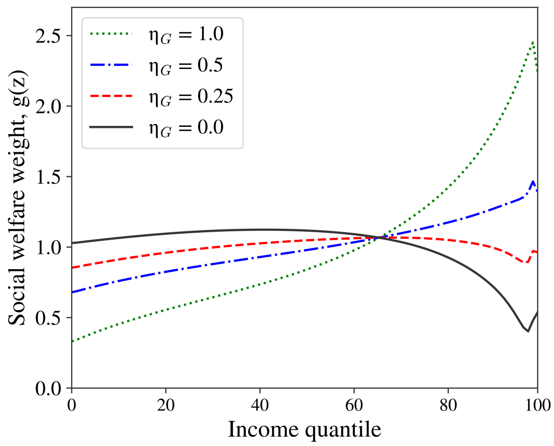

Fig. 7. Implied social welfare weights $g(z)$ from the 2019 U.S. tax system under various negative inequality externalities η_G.

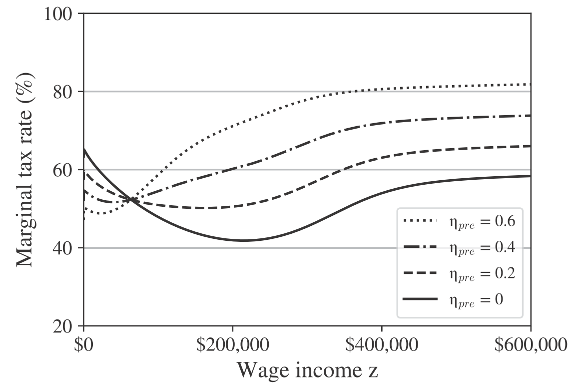

Fig. 8. Optimal income tax rates with a pre-tax income inequality externality.
Notes: The social planner is Utilitarian, and the remaining specification is identical to Fig. 5.

$\eta_G = 0.5$, and even more so for $\eta_G = 1.0$. For $\eta_G = 1.0$, the social planner values one dollar at the top equally to five dollars at the bottom.[43]

This illustrates our second main finding. The current U.S. tax schedule cannot accommodate both a socially progressive transfer motive and be significantly concerned with inequality's societal effects. The social planner may have progressive $g(z)$, implying that the government prefers to transfer one dollar from the poor to the rich ceteris paribus (as in Lockwood and Weinzierl (2016), and Hendren (2020)). The social planner may also have $\eta_G \geq 0.25$, implying a negative inequality externality of a potentially sizable magnitude. However, it cannot have both. The inequality aversion in the system as a whole is simply too small for this to be the case. It should be noted that this is, again, subject to our assumptions – particularly relevant here are the assumption of optimal tax design, Utilitarianism (Weinzierl, 2014), and the absence of migration responses (Lehmann et al., 2014).

The U.S. social planner may also have a *positive* inequality externality in mind. An inequality externality focusing on positive benefits from top-incomes could explain the puzzle of increasing SWWs at the top from Hendren (2020) (a result which is also visible in Fig. 7). If the social planner believes top-income inequality is strongly beneficial for society – through increasing innovation, economic growth, or charitable giving, for example – the implied SWWs may still be everywhere decreasing. We illustrate this graphically in Figure A7.

Several other conclusions from the inverse optimal tax literature could change if inequality externality beliefs are a salient feature of policy-making. Lockwood and Weinzierl (2016) note that TRA86 implies a substantial change in SWWs over a short time period, which could be resolved if TRA86 instead represented a change in the *inequality externality belief* of the social planner – beliefs that are arguably more malleable than the SWWs themselves. Both Lockwood and Weinzierl (2016) and Hendren (2020) also create welfare estimates that depend on inequality not being an externality (or having been considered an externality in the tax design process).[44] More generally, the inverse-optimum literature is an example of a welfare-based framework that is relatively fragile to the inclusion of an inequality externality.

5.6. Other types of inequality externalities

The preceding sections have discussed a *post-tax income* inequality externality. While such an externality could be reasonable through several motivations – some of which we outline in Section 6.1 – there is no *a priori* reason to exclude the possibility of other inequality externalities. Here we consider how the theoretical intuition changes with different types of inequality externalities in the optimal non-linear income taxation problem. Note that the optimal marginal tax formula with a post-tax income inequality externality from (5) can be written as,

$$\tau(z) = \frac{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) - \bar{G}(z)}{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) + \alpha(z)\epsilon(z) - \bar{G}(z)}. \quad (11)$$

Pre-tax income inequality externality. A pre-tax income inequality externality implies different equality impacts of the behavioral and mechanical effects. To start with the behavioral responses, note that any behavioral shift that follows from a tax increase would lead to a larger pre-tax income reduction than post-tax income reduction; pre-tax income being reduced by one unit reduces post-tax income by only $1 - \tau(z)$ units, which is between zero and one (excluding the extreme case of negative marginal rates). As such the effect of any behavioral response on pre-tax income inequality is generally larger than that on post-tax income inequality. Subsequently the pre-tax externality is more heavily affected by this channel than we saw in the post-tax case.

The mechanical effect, meanwhile, no longer has any impact on the externality. This follows from pre-tax income inequality being unchanged by the mechanical (post-tax) redistribution of income from those above the perturbation.

The optimal income tax rates in this case are

$$\tau(z) = \frac{1 + \eta_{pre} \cdot \kappa(z)\alpha(z)\epsilon(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) - \bar{G}(z)},$$

where η_{pre} is the pre-tax income inequality externality magnitude.[45] The full derivation is in Appendix I.I.

This result implies that a pre-tax income inequality externality could lead to a progressive modification of the standard Mirrlees tax rates

[43] For $\eta = 2.0$ we find negative SWWs at the bottom, indicating that the social planner would want to remove income at the bottom if this did not also increase inequality itself.

[44] Lockwood and Weinzierl (2016) calculate the welfare cost of the inequality in income growth between 1980 and 2010 as 4.3% of total economic growth in the period. Similarly, Hendren (2020) creates a preference ordering of countries' income distributions based on implied SWWs. Two parts of these calculations would be affected by an inequality externality. First, the implied SWWs from the inverse-optimum method would change under an inequality externality, as shown in this section. Second, the total welfare implications of income changes would be affected by an inequality externality.

[45] There is a subtle point to be made here about the magnitude of η_{pre}. Pre-tax income inequality is generally higher than post-tax income inequality, which influences the shadow price of each unit of inequality and hence η. To keep externality sizes similar we thus use a lower set of η_{pre} in Fig. 8 than the corresponding η_G in Fig. 5.

(where we mean progressive in the traditional sense; marginal tax rates which increase with income). We see this in Fig. 8, which shows negative pre-tax inequality externalities in the Utilitarian framework with the same specifications as in our main specification. Bottom tax rates are lower and top tax rates are higher than in the no-externality case, which is a general finding under separability. The marginal tax rates increase from 47% at the bottom to 85% at the top when $\eta_{pre} = 0.6$.[46]

Interestingly, the pre-tax income inequality externality almost removes the well-known U-shape of optimal marginal tax rates from the classical literature. Instead, the marginal tax rates generally increase in income. Compared to the classical literature (or the case of a post-tax income inequality externality), this new optimal marginal income tax schedule is closer to that observed in most developed countries. One might wonder whether governments have, to some extent, considered pre-tax inequality as an ill in itself when designing tax schedules. If so, this could explain some of the differences between the numerical simulations from optimal tax theory and real-world tax schedules.

Utility inequality externality. When considering a utility inequality externality, the behavioral channel no longer has an inequality impact. This follows from a miniscule tax perturbation from the optimum only leading to second-order utility effects. The mechanical effect would function similarly as in the post-tax income inequality case, as increasing the marginal tax rate reduces utility inequality by lowering the utility of those above the tax bracket.[47]

The optimal income tax rates in such a case are

$$\tau(z) = \frac{1 + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)},$$

where η_U is the utility inequality externality magnitude. The full derivation is in Appendix I.II. Assuming that negative weights are acceptable, using the modified SWWs $\bar{G}'(z) = \bar{G}(z) - \eta_U \cdot \bar{\kappa}(z)$ allows this result to be simplified to the standard Mirrlees case without the need for empirical variables in the modified income-based welfare weights.[48] Further, this result can be approximated in the mechanism design case through utility-based SWWs, unlike both the pre-tax and post-tax externality results.

Simply put, a utility inequality externality brings the problem closer to the standard no-externality case. Specifically, the utility problem can often be approximated by changing the inequality aversion of the SWF in the traditional (Atkinson, 1970) sense.[49] This is because the net effect of the utility inequality externality is to change the social benefit of each individuals' utility, which can be achieved through simply changing the standard SWWs.[50]

Table 2 summarizes these results.

6. Further theoretical discussion

We now turn to the more general implications of an inequality externality. The reframing of inequality as an externality leads to several simple conclusions:

- Equality itself becomes policy-relevant and has an associated shadow price.[51] The trade-off between income maximization at the bottom and the preferred inequality level becomes relevant.
- Introducing an inequality externality presents an efficiency-based reason for the state to distributionally interfere in otherwise well-functioning markets.
- A Rawlsian min–max is not the most inequality-averse modeling exercise in welfare modeling. Similarly, a Utilitarian SWF is not the least inequality-averse modeling exercise if one restricts oneself to non-increasing SWWs.
- A change in marginal tax rates can lead to a "double dividend" of both more revenue *and* an inequality level closer to what is considered optimal, both of which are welfare-relevant.
- The marginal social welfare of income at the top can be negative (Carlsson et al., 2005). In a utilitarian framework with homogeneous agents and a negative inequality externality, the total welfare effect of additional income at the top is:

$$\frac{d \sum_j g_j U(x_j, \bar{\theta})}{d x_i} = g_i \frac{\partial U(x_i, \bar{\theta})}{\partial x_i} + \sum_j g_j \frac{\partial U(x_j, \bar{\theta})}{\partial \bar{\theta}} \frac{\partial \bar{\theta}}{\partial x_i}$$

  The second term on the right-hand side comes from the inequality externality and can have significant magnitudes, as the results in Section 4 showed. The total effect depends on the relative importance of equality and consumption, a version of the familiar equity-efficiency trade-off.

This last result is particularly notable in the context of concentrated income gains. Extremely concentrated income gains – which are potentially becoming more prevalent with globalization and technical progress – are unambiguously good in standard models. The few agents receiving the additional income increase their utility, while every other agent's utility remains the same. If increased income inequality changes society, however, the other agents may be affected, positively or negatively, despite constant income levels. This is captured by an inequality externality, which illustrates a potential ambiguity in such cases. See Appendix B for further discussion.

6.1. Micro-foundations

Generally, very few assumptions are needed for an inequality externality to exist. Several different channels can be directly created from simple and mechanical microfoundations that do not rely on agents' emotional reactions, as we show in the following three simplistic examples[52]:

- Political polarization: Assume that political opinions O_i are a linearly increasing function of individual income x_i and no other factors (for simplicity). Political polarization, denoted as $\bar{P} = \varphi(\boldsymbol{O})$, is defined as an increasing function of a distributional metric φ of all opinions in the population $\boldsymbol{O}$. We assume that $\bar{P}$ enters into the individual's utility function $U_i(x_i, \bar{P}, \ldots)$. If income inequality increases, differences of opinion within the population mechanically increase as well, generally increasing $\bar{P}$ and affecting $U_i(\ldots)$. Thus, inequality leads to more pronounced political polarization and subsequent individual utility impacts.[53]

[46] This corresponds roughly to $\eta_G = 2.0$ in Fig. 5.

[47] This is more complicated outside the simple quasi-linear case, see Appendix I.II.

[48] To the extent that η_U is not an empirical variable, of course. A similar modification can be made to the income-based welfare weights in the post-tax income inequality case. However, there $\bar{G}''(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z) - \bar{G}(z)$, indicating that the modified welfare weights are dependent on $\alpha(z)$ and $\epsilon(z)$.

[49] The exception is when separability does not hold such that individuals' behavior is directly affected by the externality.

[50] There is a notable complication to this problem, namely that utility has to be carefully defined. Standard inequality metrics, such as those discussed in the post-tax income case, would not remain the same through monotonic transformations of utility. This complicates the problem both philosophically and analytically. The natural simplification we have used above is a quasi-linear utility function, in which case income changes have a one-to-one relationship with utility changes.

[51] This shadow price corresponds to η in (5) and γ in (41).

[52] An overbar indicates a society-wide variable. Bold indicates a population-sized vector.

[53] The same argument also holds for diversity of opinions more generally. A different perspective is that increased income inequality could lead to a broader diversity of opinions, carrying a positive utility impact.

Table 2
Optimal income taxation effects of various inequality externalities.

	Mechanical effect	Behavioral responses	Optimal tax rates $\tau(z)$
Post-tax income inequality	✓	✓	$\frac{1+\eta\alpha(z)\epsilon(z)\kappa(z)+\eta\bar{\kappa}(z)-\bar{G}(z)}{1+\eta\alpha(z)\epsilon(z)\kappa(z)+\eta\bar{\kappa}(z)+\alpha(z)\epsilon(z)-\bar{G}(z)}$
Pre-tax income inequality	–	✓ (stronger)	$\frac{1+\eta_{pre}\kappa(z)\alpha(z)\epsilon(z)-\bar{G}(z)}{1+\alpha(z)\epsilon(z)-\bar{G}(z)}$
Utility inequality	✓	–	$\frac{1+\eta_U\cdot\bar{\kappa}(z)-\bar{G}(z)}{1+\alpha(z)\epsilon(z)+\eta_U\cdot\bar{\kappa}(z)-\bar{G}(z)}$

Note: The table describes how each type of inequality externality functions in the optimal income taxation framework.

- Innovation/Economic growth: Assume that agents view high inequality as an incentive to work such that l_i and thus x_i are increasing functions of income inequality $\bar{\theta} = I(\boldsymbol{x})$. If so, utility can be written as $U_i(x_i(\bar{\theta}), l_i(\bar{\theta}), \ldots)$ and inequality is an externality. Further, assume that there exists some societal variable which is positively increasing in total labor supply, such as economic growth rates $\bar{g}$ or innovation levels $\bar{L}$. If this variable has an independent effect on either individual utility $U_i(\ldots)$ or productivity n_i, then income inequality has an additional welfare-relevant externality effect through $\bar{g}$ and/or $\bar{L}$.
- Income-sensitive taste for public goods: Consider the funding required for a public good project to be undertaken as $\widetilde{Q}_j$. Individual utility is defined as $U_i(x_i, \sum_j p_{i,j}q_{i,j}, \ldots)$, where the individual-specific quantity of public good j is $q_{i,j}$. Assume further that either the quantity $q_{i,j}$ or the taste variable $p_{i,j}$ varies with income levels x_i. As an example, a new youth center may be most beneficial for low-income earners, whereas an expensive opera house could be preferred by high-income earners. If income inequality $\bar{\theta}$ increases, there is less agreement on which public goods to fund and fewer projects reach $\widetilde{Q}_j$. Larger income differences in this context leads to fewer completed public projects and lower individual utility in more unequal societies.

The above examples illustrate that inequality externality channels can be mechanical in nature and can exist under only mild assumptions.[54] We also create micro-foundations for inequality effects on trust, crime, and political capture in Appendix J. Before we move on, we note that these channels may imply cascading effects. For instance, increasing political polarization could increase crime rates and hamper economic activity. We present one specific case of such secondary effects;

- Social unrest: Assume that one of the channels discussed above decreases the utility of a subset of individuals. These individuals might then prefer a high fixed cost of social unrest to living in a society with high economic inequality. If these events affect the utility of all individuals, inequality can lead to individual utility losses even for agents who were not initially negatively affected by the inequality externality. On this point we note that high economic inequality commonly precedes notable social uprisings; the French Revolution, the Russian Revolution, and the Arab Spring are some examples.

This last point illustrates that the impacts of inequality externality effects can be starkly discontinuous. In such events the externality itself would have complex optimal policy consequences as a low-probability, high-impact catastrophe event in the vein of Weitzman (2009).

6.2. Consequences in the literature

Given that the inequality externality is harder to ignore than many other externalities, a natural question is how other optimal policy models would be affected by the inclusion of an inequality externality. While this is too large of a question to fully answer in this paper, we present a few thoughts below.

First, our results question the external validity of models which rely on utility functions that only take into account individuals' income and work hours in large-scale settings. This is particularly true for numerical solutions in models focusing on inequality-related issues. As a recent example of how policy discussion can be modified through the introduction of an inequality externality we examine the model in Heathcote et al. (2020), the 2019 *EEA Presidential Address* titled "*How should tax progressivity respond to rising income inequality?*". The work analyses an optimal taxation model in a general equilibrium framework where the main benefit of higher progressivity is as insurance for idiosyncratic shocks. The authors find that tax progressivity should remain approximately unchanged given rising U.S. inequality levels, a result which is robust in both a Rawlsian and Utilitarian framework. Introducing an inequality externality would likely affect these results. Following our results (which admittedly come from a simpler model), a negative (positive) inequality externality would likely yield a more progressive (regressive) optimal tax rate when income inequality increases. The methodology in Heathcote et al. (2020) is relatively standard, and similar models are common in the economic literature. In general, we believe it would be prudent to check such results for robustness in the face of various inequality externalities or to mention the no-externality assumption explicitly.

Second, theoretical models focusing on the trade-offs between different forms of taxation such as Guvenen et al. (2019) and Jacquet and Lehmann (2021) could also be affected by an inequality externality. With an inequality externality the social planner has an added incentive to set the inequality level itself, which may be easier with one instrument or the other. Take the example of wealth taxation versus capital income taxation in Guvenen et al. (2019), where one instrument taxes a stock and the other a flow – if the externality itself is more dependent on either the stock or the flow, the relevant trade-off could be modified.

Third, cost–benefit analysis-type results that depend on income-based SWWs may be fragile to the inclusion of an inequality externality. If an inequality externality is not explicitly taken into account through either modified SWWs or through a cost estimate of income inequality itself, these frameworks implicitly assume that income inequality itself has no effect on society.

7. Conclusion

This paper has assumed that economic inequality has societal consequences and examined how welfare-based models in the economic literature are affected. To model inequality's consequences we have introduced the concept of an *inequality externality*, particularly focusing on an *income* inequality externality in the Mirrlees (1971) optimal income taxation model.

The inequality externality presents a simple way to include economic inequality's societal consequences in welfare-based models. It combines two previously disconnected subjects, both of which have

[54] Three qualifications should be noted here. First, it is not self-evident which types of inequality (income, wealth, status...) and which domains (neighborhood, country, global...) are relevant, nor which effects are likely to be large on which agents. For this paper we do not go beyond some illustrative calculations in fairly simple cases. Second, the transmission of some inequality effects are clear, such as the effect of inequality on the provision of public goods, while others are dependent on social context or perceived inequality. This implies that inequality effects can differ across societies that are equally unequal. Third, some effects are time-dependent: although not well-captured in single-period models, the basic argument remains the same.

received significant attention; inequality's societal consequences and the economic theory on externalities. The concept itself is tractable and does not assume a direction to the externality, can include other-regarding preferences but does not require them, and could be applied to various types of inequality (pre-tax income, post-tax income, wealth...). To illustrate that income inequality could be an externality we showed numerous microfounded examples that did not rely on individuals' feelings or knowledge of the income distribution. We discussed which economic trade-offs could be particularly affected by an inequality externality, and noted that the largest impact is likely when inequality itself varies significantly across policy outcomes.

In the Mirrlees (1971) optimal income taxation model, the externality introduces an additional incentive to reduce income inequality. Given that policy makers believe that income inequality itself is concerning, the analysis presented here thus recommends more inequality-reducing tax rates than those previously suggested by Saez (2001), Piketty et al. (2014), and others. We present two main new insights to the optimal income taxation literature, both of which are relevant for tax design.

First: Inequality's externality properties may have larger optimal top tax rate implications than standard tax revenue concerns. This is because raising top income tax rates reduces inequality both mechanically and through agents' behavioral responses. This contrasts to the standard revenue case, where these two effects are always opposed – there is mechanically more revenue to redistribute, but at the same time agents' behavioral responses leads to less tax revenue. This theoretical finding is borne through in numerical simulations; we observe both very high top marginal tax rates (above 90%) when inequality is a significant social bad and very low optimal top tax rates (<0%) when inequality is a social good. Our baseline estimate, to which we attach considerable uncertainty, is an 81% optimal top marginal tax rate. We thus find theoretical support for several policy arguments previously unsupported by classic economic theory, notably the high top marginal tax rates in the post-war United States and United Kingdom. The findings also imply that different beliefs about the magnitude of the inequality externality could be a potential source of political disagreement.

Second: The inequality aversion implied by the current U.S. income tax system is insufficient to explain both progressive social welfare weights *and* a realistic concern for inequality's effects on society. While the tax system may imply a preference for progressive redistribution *or* a negative inequality externality of a substantial magnitude, it is currently not able to accommodate both objectives effectively if designed optimally under our assumptions.

Finally, we also briefly discussed how our results could have policy implications beyond optimal income taxation. Given that many economic models rely on the assumption of no externalities, the idea of considering inequality's societal effects as an externality could change economic intuition in a variety of settings. We encourage further work on the topic.

Declaration of competing interest

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2024.105139.

References

Aaberge, R., 2000. Characterizations of Lorenz curves and income distributions. Soc. Choice Welf. 17 (4), 639–653.
Alesina, A., Giuliano, P., 2011. Preferences for redistribution. Handb. Soc. Econ. 1, 93–131, Elsevier.
Aronsson, T., Johansson-Stenman, O., 2008. When the Joneses' consumption hurts: Optimal public good provision and nonlinear income taxation. J. Public Econ. 92 (5–6), 986–997.
Aronsson, T., Johansson-Stenman, O., 2018. Inequality aversion and marginal income taxation. In: Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association. Vol. 111, pp. 1–32.
Aronsson, T., Johansson-Stenman, O., 2023. Optimal Taxation and Other-Regarding Preferences. In: University of Gothenburg Working Paper Series (837).
Atkinson, A.B., 1970. On the measurement of inequality. J. Econom. Theory 2 (3), 244–263.
Auclert, A., Rognlie, M., 2018. Inequality and Aggregate Demand. National Bureau of Economic Research, (24280).
Benabou, R., 1996. Inequality and growth. Natl. Bureau Econ. Res. Macroecon. Annu. 11, 11–74.
Bergh, A., Nilsson, T., Waldenström, D., 2016. Sick of Inequality?: An Introduction To the Relationship Between Inequality and Health. Edward Elgar Publishing.
Bergolo, M., Burdin, G., Burone, S., Rosa, M. De, Giaccobasso, M., Leites, M., 2022. Dissecting inequality-averse preferences. J. Econ. Behav. Organ. 200, 782–802.
Bourguignon, F., Spadaro, A., 2012. Tax-benefit revealed social preferences. J. Econ. Inequal. 10, 75–108.
Bovenberg, A.L., van der Ploeg, F., 1994. Environmental policy, public finance and the labour market in a second-best world. J. Public Econ. 55 (3), 349–390.
Breza, E., Kaur, S., Shamdasani, Y., 2018. The morale effects of pay inequality. Q. J. Econ. 133 (2), 611–663.
Carlsson, F., Daruvala, D., Johansson-Stenman, O., 2005. Are people inequality-averse, or just risk-averse? Economica 72 (287), 375–396.
Cohen, M.A., Rust, R.T., Tidd, S.T., 2004. Willingness-to-pay for crime control programs. Criminology 42 (1), 89–110.
Cooper, D.J., Kagel, J., 2016. Other-Regarding Preferences. Handb. Exp. Econ. 2 (217).
Cowell, F.A., 2000. Measurement of Inequality. Handb. Income Distrib. 1, 87–166.
Cremer, H., Gahvari, F., Ladoux, N., 1998. Externalities and Optimal Taxation. J. Public Econ. 70 (3), 343–364.
Diamond, P.A., 1998. Optimal Income Taxation: An Example With a U-shaped Pattern of Optimal Marginal Tax Rates. Amer. Econ. Rev. 83–95.
Donaldson, D., Weymark, J.A., 1980. A Single-Parameter Generalization of the Gini Indices of Inequality. J. Econom. Theory 22 (1), 67–86.
Feenberg, D., Coutts, E., 1993. An Introduction to the TAXSIM Model. J. Policy Anal. Manage. 12 (1), 189–194.
Fehr, D., Rau, H., Trautmann, S.T., Xu, Y., 2020. Inequality, Fairness and Social Capital. Eur. Econ. Rev. 129, 103566.
Fisman, R., Kuziemko, I., Vannutelli, S., 2021. Distributional Preferences in Larger Groups: Keeping up with the Joneses and Keeping Track of the Tails. J. Eur. Econom. Assoc. 19 (2), 1407–1438.
Gauthier, S., Laroque, G., 2009. Separability and Public Finance. J. Public Econ. 93 (11–12), 1168–1174.
Goodin, R.E., 1986. Laundering Preferences. Found. Soc. Choice Theory 75, 81–86.
Greenspan, A., 2014. Comments: National Association of Business Economists Conference. https://www.thefiscaltimes.com/Articles/2014/02/25/These-Top-Economists-All-Agree-Biggest-Problem-US-Faces.
Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo-Diaz, S., Chen, D., 2019. Use It Or Lose It: Efficiency Gains from Wealth Taxation. Working Paper 26284, National Bureau of Economic Research.
Harrendorf, S., 2018. Prospects, Problems, and Pitfalls in Comparative Analyses of Criminal Justice Data. Crime Justice 47 (1), 159–207.
Harsanyi, J.C., 1977. Morality and the Theory of Rational Behavior. Soc. Res. 44 (4), 623–656.
Heathcote, J., Storesletten, K., Violante, G.L., 2020. Presidential address 2019: How should tax progressivity respond to rising income inequality? J. Eur. Econom. Assoc. 18 (6), 2715–2754.
Hendren, N., 2020. Measuring Economic Efficiency using Inverse-Optimum Weights. J. Public Econ. 187, 104198.
Jacobs, B., De Mooij, R.A., 2015. Pigou meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-best Pigouvian Tax. J. Environ. Econ. Manage. 71, 90–108.
Jacobs, B., Jongen, E.L., Zoutman, F.T., 2017. Revealed Social Preferences of Dutch Political Parties. J. Public Econ. 156, 81–100.
Jacquet, L., Lehmann, E., 2021. How To Tax Different Incomes?. In: CEPR Discussion Paper Series (16571), IZA Institute of Labor Economics.
Jacquet, L., Lehmann, E., Van der Linden, B., 2013. Optimal Redistributive Taxation with Both Extensive and Intensive Responses. J. Econom. Theory 148 (5), 1770–1805.
Jones, C.I., 2022. Taxing Top Incomes in a World of Ideas. J. Polit. Econ. 130 (9), 2227–2274.

Kanbur, R., Keen, M., Tuomala, M., 1994. Optimal Non-Linear Income Taxation for the Alleviation of Income-Poverty. Eur. Econ. Rev. 38 (8), 1613–1632.

Kanbur, R., Tuomala, M., 2013. Relativity, Inequality, and Optimal Nonlinear Income Taxation. Internat. Econom. Rev. 54 (4), 1199–1217.

Kaplow, L., 2010. The Theory of Taxation and Public Economics. Princeton University Press.

Lehmann, E., Simula, L., Trannoy, A., 2014. Tax Me If You Can! Optimal Nonlinear Income Tax Between Competing Governments. Q. J. Econ. 129 (4), 1995–2030.

Lobeck, M., Støstad, M.N., 2023. The Consequences of Inequality: Beliefs and Redistributive Preferences. CESifo Working Paper No. 10710.

Lockwood, B.B., Weinzierl, M., 2016. Positive and Normative Judgments Implicit in US Tax Policy, and the Costs of Unequal Growth and Recessions. J. Monetary Econ. 77, 30–47.

Mian, A.R., Straub, L., Sufi, A., 2020. The Saving Glut of the Rich. Working Paper 26941.

Mirrlees, J.A., 1971. An Exploration in the Theory of Optimum Income Taxation. Rev. Econ. Stud. 38 (2), 175–208.

Obama, B., 2011. Remarks by the president on the economy in osawatomie, kansas. https://obamawhitehouse.archives.gov/the-press-office/2011/12/06/remarks-president-economy-osawatomie-kansas.

Oswald, A.J., 1983. Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation. J. Public Econ. 20 (1), 77–87.

Piketty, T., Saez, E., Stantcheva, S., 2014. Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. Am. Econ. J.: Econ. Policy 6 (1), 230–271.

Piketty, T., Saez, E., Zucman, G., 2018. Distributional National Accounts: Methods and Estimates for the United States. Q. J. Econ. 133 (2), 553–609.

Pirttilä, J., Tuomala, M., 1997. Income Tax, Commodity Tax and Environmental Policy. Int. Tax Public Finance 4, 379–393.

Plato, 2016. The Laws of Plato. CreateSpace Independent Publishing Platform, Estimated written 360 B.C.. Translated by Benjamin Jowett.

Rueda, D., Stegmueller, D., 2016. The Externalities of Inequality: Fear of Crime and Preferences for Redistribution in Western Europe. Am. J. Political Sci. 60 (2), 472–489.

Rufrancos, H., Power, M., Pickett, K.E., Wilkinson, R., 2013. Income inequality and crime: A review and explanation of the time series evidence. Sociol. Criminol.-Open Access.

Sadka, E., 1976. On Income Distribution, Incentive Effects and Optimal Income Taxation. Rev. Econ. Stud. 43 (2), 261–267.

Saez, E., 2001. Using Elasticities to Derive Optimal Income Tax Rates. Rev. Econ. Stud. 68 (1), 205–229.

Saez, E., 2010. Do Taxpayers Bunch at Kink Points? Am. Econ. J.: Econ. Policy 2 (3), 180–212.

Saez, E., Stantcheva, S., 2016. Generalized Social Marginal Welfare Weights for Optimal Tax Theory. Amer. Econ. Rev. 106 (1), 24–45.

Sandmo, A., 1975. Optimal Taxation in the Presence of Externalities. Swed. J. Econ. 86–98.

Seade, J.K., 1977. On the Shape of Optimal Tax Schedules. J. Public Econ. 7 (2), 203–235.

Sen, A., 1976. Real National Income. Rev. Econ. Stud. 43 (1), 19–39.

Simula, L., Trannoy, A., 2022. Gini and optimal income taxation by rank. Am. Econ. J.: Econ. Policy 14 (3), 352–379.

Stiglitz, J.E., 1982. Self-selection and Pareto Efficient Taxation. J. Public Econ. 17 (2), 213–240.

Thurow, L.C., 1971. The Income Distribution as a Pure Public Good. Q. J. Econ. 327–336.

Tsyvinski, A., Werquin, N., 2017. Generalized Compensation Principle. Technical report, National Bureau of Economic Research.

Weinzierl, M., 2014. The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice. J. Public Econ. 118, 128–142.

Weitzman, M.L., 2009. On Modeling and Interpreting the Economics of Catastrophic Climate Change. Rev. Econ. Stat. 91 (1), 1–19.

Xu, Y., Marandola, G., 2022. The (Negative) Effects of inequality on Social Capital. J. Econ. Surv..

Further reading

Jacobs, B., 2018. The Marginal Cost of Public Funds is One at the Optimal Tax System. Int. Tax Public Finance 25, 883–912.

Kelly, M., 2000. Inequality and Crime. Rev. Econ. Stat 82 (4), 530–539.

Laffer, A.B., 2004. The Laffer Curve: Past, Present, and Future. Backgrounder 1765, 1–16.

Lollivier, S., Rochet, J.-C., 1983. Bunching and Second-Order Conditions: A Note on Optimal Tax Theory. J. Econom. Theory 31 (2), 392–400.

Mankiw, N.G., Weinzierl, M., Yagan, D., 2009. Optimal Taxation in Theory and Practice. J. Econ. Perspect. 23 (4), 147–174.

Manning, A., 2015. Top Rate of Income Tax. Centre for Economic Performance's Election Analysis.

Masterman, C.J., Viscusi, W.K., 2018. The Income Elasticity of Global Values of a Statistical Life: Stated Preference Evidence. J. Benefit-Cost Anal. 9 (3), 407–434.

Saez, E., Slemrod, J., Giertz, S.H., 2012. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. J. Econ. Lit. 50 (1), 3–50.

Salanie, B., 2011. The Economics of Taxation. MIT Press.

Schmidt, U., Wichardt, P.C., 2019. Inequity Aversion, Welfare Measurement and the Gini Index. Soc. Choice Welf. 52 (3), 585–588.

Tuomala, M., 2016. Optimal Redistributive Taxation. Oxford University Press.