Inequality as an Externality: Consequences for Tax Design[*]

Morten Nyborg Støstad[†] and Frank Cowell[‡]

Abstract

This paper proposes to treat income inequality as an economic externality in order to introduce the societal effects of inequality into welfarist models. We introduce such effects in a simple and generalizable welfarist framework and show that they can have sizeable optimal policy consequences that cannot be captured by standard risk aversion or social welfare weights. Novel policy implications are illustrated through the classical optimal non-linear income taxation model, where the social planner must face a trade-off between collecting revenue and changing income inequality levels. Resulting policy consequences are disproportionately located at the top, where optimal marginal tax rates are strongly and robustly dependent on the magnitude of the inequality externality. We use several real-world examples to show that tax policy previously unsupported by optimal taxation theory can be explained in our framework. The findings indicate that the magnitude of the inequality externality could be considered a crucial economic variable. *JEL* Codes: H21, H23, D62, D63

[*] We thank Stéphane Gauthier, Marc Fleurbaey, Daniel Waldenström, Olof Johansson-Stenman, Fredrik Carlsson and Karine Nyborg for helpful comments and discussions. We have also benefited from suggestions from Etienne Lehmann, Marie Young Brun, Max Lobeck, Elif Cansu Akoğuz, Stefanie Stantcheva, Thomas Blanchet, Antoine Bozio, François Fontaine, Thomas Piketty, and seminar participants at ECINEQ 2019, the Paris School of Economics, the GT Économie de la Fiscalité, and the University of Oslo. Version: June 15, 2021.

[†] Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Phone: +33766142152. Email: morten.stostad@psemail.eu (corresponding author).

[‡] London School of Economics, Houghton Street, London, W2CA 2AE, UK. Email: f.cowell@lse.ac.uk.

I Introduction

Economists, policy makers and philosophers have long argued that an unequal distribution of resources affects society and thus individuals. How this happens is up for debate. An incomplete set of potential channels include inequality's effects on economic growth, interpersonal trust, political polarization, social stability, innovation, crime, or individual health. While there is no current consensus on which of these potential effects are most relevant, nor their combined welfare impact, there is a shared understanding that inequality in resources itself is policy relevant.

Despite this, and despite a large empirical academic literature attempting to explore such effects,[1] income inequality itself does not affect individuals in the vast majority of economic models. While there are other consequences of income inequality in standard welfarist frameworks,[2] agents are usually assumed to be individually indifferent to changes in societal inequality as long as their own income is held constant. In other words, income inequality itself has no effect on society or the individual. This holds true in nearly all welfarist models.[3]

This paper introduces the societal effects of income inequality into a simple and generalizable welfarist framework. Further, it shows that such an *inequality externality* can have significant consequences for optimal public policy. Simply put, including even a moderately-sized inequality externality in the analysis tends to drastically change theoretical outcomes and policy advice.

To illustrate we use the example of an income inequality externality in the Mirrlees (1971) optimal income taxation (OIT) model. Standard models have only considered the *revenue* benefits of taxation; we must also consider the *equality* effects of taxation. This changes modeling outcomes in a way that cannot be captured by standard individualist parameters or appropriate utility-based social welfare weights. The novel effects are large, intuitively appealing, and reflect how both policy leaders and economists describe policy goals. The consequences are particularly striking for optimal top tax rates. We find several examples of real-world policies that cannot be rationally justified under standard optimal taxation rules, yet are welfare-optimizing when inequality is an externality.

The approach we use can be extrapolated to a wide array of economic models. The key assumption we change in the OIT framework is the widely used assumption of fully individualist utility functions. Many other modeling frameworks concerned with inequality-related issues has relied, perhaps too heavily, on the same assumption by using social welfare functions as a catch-all "inequality aversion" term. As we will show this is not sufficient if inequality in resources affects individuals. Indeed, it yields substantially different optimal policy conclusions. While this paper concerns itself solely with the OIT model, we thus question which other influential economic models

[1] See, for example, Bergh et al. (2016), Rufrancos et al. (2013), Fairbrother and Martin (2013), You (2014), and Cingano (2014) for empirical academic works on inequality's effects on (respectively) health outcomes, crime levels, interpersonal trust rates, corruption rates, and economic growth levels. It has also been argued that inequality could foster political polarization and the growth of populist movements in Bonica et al. (2013), Pastor and Veronesi (2018), and Burgoon et al. (2018).

[2] Most often (i) the cumulative effect of diminishing marginal utilities of income, and (ii) inequality-averse social welfare functions, see Section II.

[3] The few exceptions generally discuss altruism or relative income concerns. Such models are based on emotional reactions to inequality, which make their optimal policy impact debatable, and still largely neglect any societal effects of a skewed income distribution.

could be similarly impacted by an inequality externality.

We now briefly introduce why inequality could be an externality and discuss the main findings from our OIT application.

The Effects of Inequality on the Individual Individualist utility functions are generally justified by arguing that feelings of altruism or jealousy should be irrelevant for optimal policy design. We do not aim to challenge this assertion. We instead note that the potential consequences of inequality – changes to political efficiency, economic growth, interpersonal trust, innovation, crime, public good funding, social stability, and so on – imply that individuals can have a strong motive to care about resource equality even absent any other-regarding preferences. Even a perfectly self-centered individual may care a great deal about the *effects* of inequality; people do not have to be "nice", or envious, to be affected by inequality.

As an example, imagine a perfectly self-interested individual in a society where income inequality increases crime. Say that income inequality and thus crime increases, and that the individual's bike is stolen as a consequence. The individual experiences a negative shock and would undoubtedly, absent any other changes, prefer the prior (more equal) state of the world. Thus, if inequality leads to more crime, inequality should enter her utility function.[4] The equivalent argument could be made with any other variable affected by inequality.[5]

We formalize our model by analyzing income inequality itself as an economic externality which enters into the utility function. Our rationale for this externality formulation is the following. First, individual labor supply decisions affect income inequality, but workers are not incentivized to take this effect into account when making their labor supply decisions. Second, income inequality could affect the utility of any other individual even though those individuals did not choose to incur any such costs or benefits. Thus, income inequality is an externality. While our focus in this paper is on an *income* inequality externality, the broader argument can be applied to other forms of inequality, perhaps particularly in wealth.

This idea was first proposed by Thurow (1971), who argued that the income distribution could be modeled as a pure public good. In a short paper, Thurow showed that the First Welfare Theorem no longer holds in such a setting. While the empirical literature on inequality's effects has grown drastically since then, reflecting more available data and improved empirical methods, the theoretical literature has largely forgotten Thurow's idea. The two main exceptions are the other-regarding preference literature, whose issues relating to optimal policy we described above, and a brief discussion in Alesina and Giuliano (2011) on how inequality could affect individual consumption.[6]

[4] Although the inequality externality acts at least partly through individual resources in this example, it is not a necessary condition for the externality to exist. We discuss further in Section II.

[5] The argument is similar to a part of the argument as to why consumption enters the utility function. We often value what consumption will give us – such as the satisfaction after a good meal, or having good health – and not necessarily the concept of consumption itself.

[6] Other notable related concepts are the status concerns discussed in Frank (1985) and the relative income concerns explored in Clark et al. (2008).

We go further and create a more extensive framework for discussing these ideas, including micro-founding a large set of potential inequality effects that do not necessarily only affect consumption. We also discuss specific policy implications. Overall, this work aims to revitalize the theoretical literature on inequality's externality effects and create a simple and generalizable framework around what we name inequality's *three welfarist consequences*; (i) unequal marginal utilities of income, (ii) the welfare-loss from differing social weights, and (iii) inequality's externality effects.

The Mirrlees (1971) Model on Optimal Income Taxation To illustrate how our framework changes classical economic theory we use the Mirrlees (1971) model. As a widely used model describing optimal income taxation (OIT), it represents both an important pillar of public economics and an appropriate example of how standard economic models rely on the no-externality assumption.

The Mirrlees model has largely focused on two equality-relevant factors; diminishing marginal utilities of income and social welfare weights. These represent, respectively, the differing value of a dollar for different agents and philosophical fairness concerns by the social planner. While many papers have explored modifications to the continuous Mirrlees model, including various types of externalities – Kanbur and Tuomala (2013) and Rothschild and Scheuer (2016) are some recent examples, exploring relative income concerns and rent-seeking respectively – income inequality itself has not been considered as an externality in the continuous model before this paper.[7] The non-linear income externality structure we employ is also novel.

The income inequality externality we introduce is mathematically and conceptually distinct from the two standard equality-relevant model characteristics we noted above. While agent behavior remains the same, the social planner's incentives for taxation changes. In general, whereas the standard Mirrlees (1971) model only focuses on the revenue effects of taxation, we must also consider the *equality effects* of taxation.

We find that the modified OIT model generates unambiguously more progressive (regressive) marginal tax rates under a negative (positive) inequality externality. The magnitude of the change is potentially large and disproportionately changes tax rates at the top of the income distribution. The two factors driving the change correspond to the equality versions of the well-known behavioral and mechanical effects of taxation from the classical literature described in Saez (2001).[8] With an inequality externality, these effects do not only change the revenue collected but also equality itself. This has a pertinent welfare effect which differs from the standard case. In particular, the behavioral response, which always decreases revenue, has a changing equality impact over the distribution and can thus be welfare positive in certain situations. The comparison to conventional results emphasizes that our model utilizes standard methods while introducing a key welfare impact

[7] Aronsson and Johansson-Stenman (2020), developed concurrently with this paper, discusses various types of other-regarding preferences including classical Fehr and Schmidt (1999) inequality aversion in a three-agent OIT model. Our paper differs in both motivation and the analytical introduction of the externality dimension, see Section III.

[8] The behavioral response describes how agents at a certain tax bracket change their work decision when their marginal tax rate increases. The mechanical effect describes the increase in tax revenue from any agent above the tax bracket, where work decisions are less impacted due to the unchanging marginal tax rates.

of taxation that leads to new policy implications.

Top marginal tax rates are particularly sensitive to the externality. There are two main reasons. First, the location of the tax-payer is crucial when evaluating equality effects, which means that the social planner's incentive to change incomes will be larger towards the ends of the distribution.[9] Second, given that only top income-earners can be specifically targeted with marginal tax rates, the equality effect induces more changes to top marginal tax rates than to bottom marginal tax rates, where changes to the marginal tax rate affect the whole distribution.

Our numerical simulations support this argument. Optimal lower- and middle-class marginal tax rates are less affected than optimal top tax rates, which change drastically. Given standard parameter values, various magnitudes of the inequality externality can lead to almost any marginal tax rate near the top of the distribution. We observe values ranging from near-zero to 90% in the top 5% of the distribution under reasonable externality values, holding all other parameters constant. We also find optimal top tax rates above the revenue-maximizing Laffer rate, as direct equality effects imply that the social planner might trade off some revenue for changed equality levels.

This builds theoretical support for previously unsupported policy arguments, such as the high post-war top marginal tax rates in the US and the UK (if inequality is a negative externality), or the low contemporary top marginal tax rates in many countries (if inequality is a positive externality and even if the social planner is Rawlsian). Other theoretical results in our model also differ from the original Mirrlees findings. The famous result that the optimal marginal tax rate for the very top agent is zero no longer holds; the social planner introduces a tax or subsidy aimed at correcting the top agent's socially incorrect work choice. We also observe negative optimal marginal tax rates. Overall, our results indicate that the magnitude and direction of the inequality externality should be considered a crucial variable when constructing optimal policy.

This paper will focus principally on two issues. First, building the theoretical case for an inequality externality. Second, deriving analytical and numerical results when taking account of an inequality externality into the OIT model. The paper is organised as follows. Section II 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 III incorporates an inequality externality in a standard OIT model and investigates the impact of the externality on optimal tax rates. Section IV concludes.

II Inequality and Social Welfare: An Externality Approach

There are several channels through which inequality can affect social welfare. This section intends to clarify how inequality has traditionally influenced welfare analysis, and to illustrate which effects have been considered through which welfare channels. We first examine what we call the *three*

[9] This contrasts to revenue effects, where the location of the tax-payer is secondary to the magnitude of potential revenue. In the extreme case, the Rawlsian min-max, "one dollar is one dollar" as long as it is not taken from the very bottom of the distribution.

welfare consequences of inequality and build a framework to describe the causes and formulations of each consequence. Then we explore the role of the inequality externality in-depth and create simple micro-foundations for certain types of inequality externalities.

II.A The three welfare consequences of economic inequality

Inequality-related concerns can enter into the formulation of social welfare comparisons through three main channels. The three are mathematically and intuitively distinct; except for special cases, they cannot be interchanged. Each channel is caused by a particular theoretical mechanism. The first two channels, the diminishing marginal utility of income and generalized social weights, have played large roles in modern economic theory. The third, the inequality externality, has not.

1 Diminishing marginal utility of income. Invoking this assumption for individual agents – which can be based on risk aversion, or simple human nature – has the following consequence. Within a utilitarian social welfare framework, income would optimally be allocated where the (social) marginal utility of consumption is highest. Assuming identical utility functions, this is at the bottom of the income distribution. The presence of income inequality runs contrary to this logic. Dalton (1920) summarizes the consequences as "the extreme wastefulness from the point of view of economic welfare of large inequalities of income". We can think of this as capturing *indirect social concern* for inequality: society is indifferent to inequality of utility, but individuals' utility is a strictly concave function of income, and therefore society is concerned about the dispersion of income in the population.

2 Income- or utility-sensitive social weights. Social weights come from the social welfare function, and represent the value society gives to an agent receiving another unit of utility. They are inherently philosophical, and are imputed from ethical principles involving distributional justice and fairness. If social weights are inversely dependent on income or utility, the social planner is incentivized to reduce inequality, although inequality has no individual cost as such. We can think of this as *direct social concern* for inequality. In view of its prevalence in the literature, this conventional OIT approach is further discussed in Appendix A.

It is clear that, through either of these two channels, inequality can be characterised as a public "bad". Despite this, inequality has no individual cost through either channel. Any individual with a given income is indifferent if inequality levels change. Moreover, any societal effects of inequality are absent. In the remainder of the paper we focus on a third channel:

3 The inequality externality. Income inequality may directly affect individual utility, an idea already found in Thurow (1971):

> *"The distribution of income itself may be an argument in an individual's utility function. This may come about because there are externalities associated with the distribution of income. Preventing crime and creating social or political stability may depend on*

> *preserving a narrow distribution of income or a distribution of income that does not have a lower tail. Alternatively, individuals may simply want to live in societies with particular distributions of income and economic power."*

Within this third channel we may identify two distinct strands. First, *other-regarding preferences* (ORP), where individual utility is directly affected by the income of others via altruism, envy, and so on. Second, *inequality effects*, including crime, political polarization, health outcomes, and more. Either strand can be seen as an economic externality. With few exceptions – most importantly Alesina and Giuliano (2011) – the latter preference-independent externality dimension has been largely theoretically neglected since Thurow.[10]

Before we discuss these two strands in detail, we note that an income inequality externality cannot be fully captured – or approximated – by a combination of the two other channels. This follows immediately from the realization that introducing an income inequality externality leads to a socially sub-optimal individual labor decision even in the absence of any taxes. Such an outcome cannot be achieved through the two other channels. We show a simple proof in the case of social weights and the inequality externality in Appendix A.I. As a result, neglecting externality issues leads to differing policy conclusions, which we discuss in the context of the OIT problem in Section III.[11]

The three channels are outlined in Table I, and further discussed in Appendix A.

Table I

The Three Welfarist Consequences of Inequality

	Diminishing marginal utility of income	Social welfare weights	Inequality externality
Formulation	$\int_i g_i U_i(\boldsymbol{x_i}, \bar{\theta}, ...)di$	$\int_i \boldsymbol{g_i} U_i(x_i, \bar{\theta}, ...)di$	$\int_i g_i U_i(x_i, \bar{\boldsymbol{\theta}}, ...)di$
Causes	The decreased value of a dollar with increased income	Societal considerations of fairness, philosophical concerns	The societal effects of inequality, other-regarding preferences

Note: The three channels through which inequality could impact welfarist modeling. For each channel the key expression is highlighted in bold.

[10] Alesina and Giuliano (2011) considers both ORP and how inequality might affect consumption and thus utility. Our paper also explicitly discusses the potential utility impact of inequality through non-consumption channels, and introduces the externality into the OIT problem. Other papers on topics related to inequality as an externality include Pauly et al. (1973) and Ashworth et al. (2002), where redistribution is modeled as a pure public good. Lindbeck (1985) discusses the consequences of inequality on macroeconomic policies, Anbarci et al. (2009) suggest an externality effect of rising inequality through an increase in traffic fatalities, and Rueda and Stegmueller (2016) consider crime as a negative externality of inequality.

[11] This is due to the externality being in terms of *income* inequality (or another parameter the individual chooses, such as wealth). If the inequality externality is in terms of utility, the individual's labor decision is socially optimal and the (utility) inequality externality functions as a change to the social weights.

II.B Channel 3 – the inequality externality

The inequality externality consists of both strands in the seminal Thurow (1971) quotation. Here we specifically discuss an income inequality externality.

1 Other-regarding preferences (ORP). Other-regarding preferences are the direct effects of other agents' income on the individual. If income inequality changes, an agent with ORP will be affected regardless of whether any of his/her own circumstances change (Cooper and Kagel, 2016). Such preferences are often described as either altruism (positive ORP) or jealousy (negative ORP). Relative income concerns are another example.

We note that ORP are not equal to stated equality preferences, which are also affected by other factors, e.g. the inequality effects we discuss below, fairness considerations, or the prospect of upwards mobility (Benabou and Ok (2001)).

2 Inequality effects. This component of the inequality externality enters indirectly into individual utility. We define an inequality effect as the channel by which inequality affects utility through a secondary variable. Inequality effects can be directly created from simple microfoundations, as we show in the following examples:

- 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 of all opinions in the population $\boldsymbol{O}$. The overbar indicates a society-wide variable. We assume that $\bar{P}$ enters into the individual's utility function $U_i(x_{i,}\bar{P}, ...)$. If income inequality $\bar{\theta} = I(\boldsymbol{x})$ increases, differences of opinion within the population mechanically increases as well, generally increasing $\bar{P}$ and affecting $U_i(...)$. Thus, inequality leads to more pronounced political polarization and subsequent individual utility impacts.[12]

- Innovation / Economic growth: Assume that agents view inequality as an incentive to work such that h_i and thus x_i are increasing functions of income inequality $\bar{\theta} = I(\boldsymbol{x})$. If so, utility can be written as $U_i(x_i(\bar{\theta}), h_i(\bar{\theta}), ...)$ and inequality is immediately 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 impact on either individual utility $U_i(...)$ or productivity n_i, then the labor choice change has an additional welfare-relevant externality effect through $\bar{g}$ and/or $\bar{L}$.

- Income-sensitive taste for public goods: Consider the funding for a public good project $\bar{Q}_j$. Individual utility is defined as $U_i(x_i, \sum_j q_{i,j}, ...)$, where individuals' expected benefits from the public good j is $q_{i,j}$. Assume further that the taste parameter $q_{i,j}$ varies with income levels x_i. As an example, a new youth center may be most beneficial for low-income earners,

[12] 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.

whereas an expensive opera house could be preferred by high-income earners. If inequality $\bar{\theta}$ increases, the average $\bar{Q}_j$ decreases 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 effects can be rather mechanical in nature and can exist under only mild assumptions.[13] However, they are not necessarily the largest effects. While the present paper does not propose an exhaustive list of all potential inequality effects, we present three other secondary effects that we believe could be particularly impactful.

- Trust: Assume that individuals have higher trust $t_{i,j}$ in other individuals who share a set of similar characteristics, where the set of relevant characteristics is denoted as the vector $\overrightarrow{T}$. If income x is part of $\overrightarrow{T}$, or causes changes in individual parameters that are, a change in income inequality $\bar{\theta}$ would decrease individual i's general trust levels $T_i = \sum_i t_{i,j}$. If T_i enters into individual utility $U(x_i, T_i, ...)$, income inequality has an indirect utility impact.

- Crime: Assume that criminal activity gains a fraction α of another agent's income x_j, subtracting a fixed risk cost. Further assume that the opportunity cost of crime is a wage-paying job with a salary proportional to the agent's income x_i, and that agents will commit crime if it is profitable. We define the Gini coefficient as $\overline{\theta}_G = \sum_i \sum_j (x_i - x_j)$. If $\overline{\theta}_G$ increases, criminal activity also increases with subsequent utility impacts. As richer individuals can spend more of their income to protect their assets, the effect might be moderated or even overturned.

- Political capture: Assume that the political process is affected by a voting procedure between discrete options $\{\overline{V}_1, ..., \overline{V}_m\}$ where each agent has a number of votes v_i proportional to their income x_i. Assume further that individual utility $U_i(x_i, \overline{V}_k, ...)$ is dependent on the outcome of this political process, with varying individual preferences. Changing income inequality $\bar{\theta}$ will mechanically change voting outcomes by giving higher-income agents a larger vote share. As the vote outcome affects the individual utility of every agent – positively or negatively – inequality indirectly affects individual utility.

These channels may imply cascading effects. For instance, decreased generalized trust could increase crime rates and hamper economic activity. We present one specific case of such tertiary effects;

- Social unrest: Assume that high polarization, significant political capture, low trust, or negative ORP regarding high incomes decreases individual utility. Following the channels described above, a subset of individuals might prefer a high fixed cost of social unrest to living

[13] 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.

in a society with extremely high inequalities. If so, such preferences can lead to revolts, revolutions, or other types of social unrest. If these events impact the utility of all individuals, inequality can lead to individual utility losses even for agents who were not negatively affected before the social unrest.

Other inequality effects could include inequality's effects on health outcomes, social stability, education levels, individual freedoms, the distribution of power, climate change, and more.

Consider how to model these effects. If we just focus on other-regarding preferences, the externality could be appropriately captured by introducing an inequality metric $\bar{\theta}$ directly into the utility function. If the externality arises principally from an effect such as innovation, then the effect comes through individual income and consumption and so could be captured by a term such as $x_i(\bar{\theta})$.[14] The other inequality effects – polarization, social unrest, and so on – generally have non-consumption utility impacts, and should thus be captured in an expression such as $\overrightarrow{\Psi}(\bar{\theta})$ where $\overrightarrow{\Psi}(.)$ is a vector of inequality effects. Putting these three together we might consider the following specification of the utility function:

$$U_i(x_i(\bar{\theta}), \bar{\theta}, \overrightarrow{\Psi}(\bar{\theta}), ...). \tag{1}$$

Detailed information on each component in the specification (1) is unlikely to be available.[15] It is also unnecessary: the separate contributions are less important than the overall impact of inequality in the utility function. This overall impact is sufficient to describe how the externality functions. So the specification (1) could be written more compactly as the simplified form:

$$U_i(x_i, \bar{\theta}, ...) \tag{2}$$

where the term $\bar{\theta}$ represents the impact on the individual of the total inequality externality.

As expressed in the form (2), the inequality externality as a whole 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. Furthermore, the inequality externality can exist with just one of the two components. For instance, one could be wholly self-serving and still have a utility function that is strongly dependent on inequality from the inequality effects component, a scenario that may be appropriate for many people.

II.C Implications

Considering income inequality as an externality immediately leads to some intuitive conclusions.

- Income equality becomes policy-relevant in itself.
- Individuals make a socially suboptimal work decision. With a negative income inequality

[14] This channel is discussed in Alesina and Giuliano (2011).

[15] We would need to know the exact health function, the exact crime function, and so on, in addition to individuals' other-regarding preferences.

externality those at the top work too much and those at the bottom work too little. The opposite is true for a positive income inequality externality.

- The focus of the optimal taxation literature has principally been on the mechanical, behavioural, and direct welfare effects of taxation. By introducing an inequality externality we must also consider the *equality effects* of taxation.

- 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:

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

The second term on the right-hand side comes from the inequality externality and can have significant magnitudes, as we will show in Section III. It is negative if inequality increases ($\frac{\partial \theta}{\partial x_i} > 0$)[16] in a society with a negative inequality externality ($\frac{\partial U(x_j,\theta)}{\partial \theta} < 0$). It can be larger than the first term (the individual benefit from the consumption increase), indicating that additional income at the top can be detrimental if inequality is sufficiently socially disruptive.[17] The total effect depends on the relative importance of equality and consumption, a version of the familiar equity-efficiency trade-off.

The last point may seem controversial. In the context of jealousy effects (ORP), Piketty and Saez (2013) argue that "hurting somebody with higher taxes for the sole satisfaction of envy seems morally wrong". In the context of inequality effects, however, the interpretation is perhaps more intuitive. Imagine, for instance, an extremely high-income agent who has a resource-determined control over the political process. If this political control hurts lower-income agents, taxation of the high-income agent designed to offset the political effects is intuitive and can be optimal in our framework. The same argument holds for other inequality effects.

This result is particularly important 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 winners or losers despite constant income levels. This is captured by an inequality externality, which illustrates the potential ambiguity in such cases. See Appendix A for further discussion.

[16] Inequality measures generally have $\frac{\partial \theta}{\partial x_i} \neq 0$ for virtually all agents. The absolute Gini coefficient, for instance, can be written as $I_{\text{Gini}}(\mathbf{x}) = \sum_{i=1}^{n} \kappa(x_i) x_i$,where the indexing of i has been chosen in increasing order of x_i, such that $\kappa(x_i) := \frac{1}{n}\left[2\frac{i}{n} - \frac{1}{n} - 1\right]$. Evidently $\frac{\partial I_{\text{Gini}}}{\partial x_i} = \kappa(x_i)$.

[17] Even though the individual's marginal effect on the inequality metric is small (of the order $\frac{1}{n}$), it being summed over n agents creates a non-negligible welfare effect on the same order of magnitude as marginal changes in consumption.

III Tax Design

In the following section we will introduce an income inequality externality into the Mirrlees (1971) OIT model.

The Mirrlees approach is the standard starting point in the optimal income taxation literature.[18] The primary cost-benefit trade-off in this model has been between revenue collection and the efficiency losses from taxation. Inequality enters the model largely through the changing individual benefits of additional private income. In other words, the model incorporates potentially changing marginal utilities of income and social welfare weights. Traditionally, however, it is assumed that post-tax income inequality has no effects on the individual.[19]

Following the model, optimal tax-policy disagreements have largely focused on three principal areas: the extent of behavioral responses to taxation, the choice of a social welfare function, and the shape of the current wage-earning ability distribution. The contribution of this work is to suggest another core dimension to the optimal income tax problem; the extent to which income inequality is an externality.

Works since at least Sandmo (1975) have examined the effect of various other externalities on the optimal tax problem. Aronsson and Johansson-Stenman (2016) and Aronsson and Johansson-Stenman (2020) are closest to our analysis. The former examines the effect of inequality aversion on income taxation, focusing on the first-best case and Pareto-optimal taxation. The latter examines ORP in the second-best model, which is mathematically related to our analysis here, although they use three discrete agent types and focus on three well-known models of social preferences rather than on the overall effects of inequality.[20] The potential for a direct focus on distributional concerns in the OIT model has also been noted by Kanbur et al. (1994) in terms of poverty concerns and Prete et al. (2016), which employs a non-welfarist approach and piecewise taxation to minimize post-tax income inequality. We differ from these two approaches by introducing the distributional term directly into the agent's utility function. Such an approach, focused on the individual's utility function, is related to the relative income OIT literature, see Boskin and Sheshinski, 1978; Oswald, 1983; Tuomala et al., 1990; Persson, 1995; Aronsson and Johansson-Stenman, 2008, 2010; Kanbur and Tuomala, 2013; Aronsson and Johansson-Stenman, 2015. This literature has generally focused on the potential negative externalities of other agents' incomes, whereas our model has a non-linear externality term dependent on the other agent's location in the post-tax income distribution.

A recent literature on rent-seeking is also conceptually related to this paper through the externality dimension. Piketty et al. (2014) introduces tax avoidance and compensation bargaining into the standard model and establishes the relevant elasticities in the case of such externality-inducing behavior, focusing particularly on top income taxation. Rothschild and Scheuer (2016) explores a

[18] See for example Diamond and Mirrlees (1971); Atkinson and Stiglitz (1976); Mirrlees (1976); Diamond (1998); and Saez (2001). Non-analytical solutions to the standard problem are found in Blundell and Shephard (2011) and Aaberge and Colombino (2013).

[19] Pre-tax income inequality is only important for the individual insofar as it can increase tax revenue. In this context, higher pre-tax incomes anywhere are a positive externality.

[20] Their models are based on Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Charness and Rabin (2002).

model with a traditional sector and a rent-seeking sector, where the social planner must correct for the rent-seeking externalities without directly observing the sector difference. Lockwood et al. (2017) considers the allocation of talented individuals under the assumption that productivity externalities range from positive to negative from low-paying to high-paying jobs. Both of these latter papers include a dimension of imperfect targeting, unlike this work, which dampens the externality benefit of income taxation and changes the scope of the analysis. In general, our work differs from the rent-seeking literature by considering post-tax income differences *themselves* as a negative externality, regardless of origin. This naturally increases the role of income taxation in the optimal policy solution.

This paper is the first work to focus on the potential issue of inequality as an externality in the OIT model. It is also, to the best of our knowledge, the first to use a continuum of agents in the OIT problem when agents have non-linear and comprehensive preferences for other agents' incomes.[21]

We consider the second-best solution for a non-linear optimal income taxation schedule in the presence of a post-tax income (consumption) inequality externality. The externality is modeled by introducing a post-tax income inequality term in the individual's utility function such that utility is determined by $U(x, h, \bar{\theta})$ where x is consumption, h is hours worked, and $\bar{\theta}$ is some post-tax income inequality metric. We solve the problem with both the small perturbations method of Saez (2001) and under a full analytical specification.

1 Inequality's effect on individual utility Post-tax income inequality $\bar{\theta}$ is a society-wide parameter, indicated by the overbar, which is determined as a function of all agents' post-tax income.[22] Agents do not take their own effect on income inequality into account when making labor decisions, as their effect on the inequality metric is on the order of $\frac{1}{n}$ and thus negligible with large n.[23] However, their actions have welfare-pertinent effects as the change in income inequality impacts n other agents.

Instead of fully specifying a functional form of individual utility we suggest to use the marginal rate of substitution between post-tax income inequality and individual income, $\eta_i = MRS_{x_i\bar{\theta}} = -\frac{dU_i/d\bar{\theta}}{dU_i/dx_i}$. This η_i measures how much consumption the individual would give up for or pay for one unit decrease in the relevant inequality metric. If $\eta_i = 0 \,\forall\, i$ we return to the standard case. For the main specification we set η to be constant for all agents and income levels. This corresponds to a homogeneous inequality externality and assumes that the (absolute) inequality metric affects utility proportionately to how consumption affects utility. This definition implicitly assumes separability in inequality and income; if separability is maintained, individuals' work decisions do not change

[21] Aronsson and Johansson-Stenman (2020) solves the second-best problem for a three-agent model where agents have preferences as described in Fehr and Schmidt (1999).

[22] The analytical problem changes with different types of inequality, e.g. pre-tax income inequality or utility inequality. Post-tax income inequality is our main focus, as it is the metric we believe is most likely to have the inequality effects we discuss in Section II.B.

[23] As we use a continuum of agents, this effect is indeed negligible in our model. Furthermore, the assumption is theoretically supported by Dufwenberg et al. (2011), which finds that individuals' demands are independent of other allocations given a separability condition that is satisfied here. Due to this assumption, it is not necessary for the individual to be aware of or estimate the magnitude of the income inequality externality.

based on the inequality level and heterogeneous inequality externalities could easily be introduced (as the net social welfare weight of the externality, determined by $\bar{\eta} = \int_i g_i \eta_i di / \int_i g_i di$ where g_i is the social welfare weight, would be the policy-determining variable). We note that in a quasi-linear utility function, which we use in the small perturbation solution in Appendix B, these assumptions imply an additive linear income inequality externality.[24] The more general form is found in the analytical solution of the problem in Appendix C.

To complete the model we need a post-tax income inequality metric $\bar{\theta}$. For the main specification we use a particular form of the (absolute) Gini coefficient in post-tax income, which has the simple form:

$$\bar{\theta}_{\text{Gini}}(\boldsymbol{x}, F) = \int_{\underline{z}}^{\bar{z}} \kappa_G(z) x(z) dF(z), \tag{3}$$

where x is after-tax income (consumption), z is total individual earnings, and

$$\kappa_G(z) = 2F(z) - 1 \tag{4}$$

is the weight of the agent in the Gini (Cowell, 2000). This weight only depends on the *rank* of the individual in the distribution, which is equal in the pre- and post-tax Gini if the second-order conditions hold (which we assume). In other words, we have that $F(z) = F(x)$ which allows us to write the Gini as in Equation 3.[25] Expression 3 shows that the absolute Gini can be calculated as a sum of weighted incomes in the population, where the weight $\kappa_G(z)$ depends only on the rank of the agent in the pre-tax income distribution.

This specification streamlines the analytical problem. One can also use other inequality metrics based on other types of rank-specific weights $\kappa(z)$ where $\int_0^\infty \kappa(z) dF(z) = 0$, such as those in the Lorenz (Aaberge, 2000) or S-Gini families (Donaldson and Weymark, 1980). In the main text we also perform a robustness test with a generalised Gini with weights of the following form:

$$\kappa_T(z) = (q+1)F(z)^q - 1. \tag{5}$$

The Gini corresponds to $q = 1$, while larger q approximates top income share inequality metrics (while remaining analytically tractable). This addresses the issue that the Gini can over-weight middle incomes. We further extend our analysis to the S-Gini in Appendix D.II. Absolute inequality metrics are used to keep scale invariance.

III.A The Optimal Marginal Tax Schedule

To calculate the optimal income taxation results we use the small perturbations method from Saez (2001).

[24] Utility is thus a monotonic transformation of $U(x, h, \bar{\theta}) = log(x - v(h) - \eta\bar{\theta})$. Conceptually, the inequality externality does not have to be linear. If the externality is squared such that the term in the utility function is $\eta(\bar{\theta} - \theta_{opt})^2$, the MRS becomes $2\eta(\bar{\theta} - \theta_{opt})$ which is dependent on the distance from the optimal inequality level θ_{opt} (see Appendix D.III). One could also imagine an inequality externality dependent on the current income level, which would change agent behavior with potentially complex consequences. We leave this to be explored by later works.

[25] The original expression from Cowell (2000) is $\int_{\underline{x}}^{\bar{x}} \kappa(x) \cdot x \cdot dF(x)$.

The resulting marginal tax rates $\tau(z)$ at earnings z are (see Appendix B for full derivation),

$$\tau(z) = \frac{1 + \Upsilon(z) - \bar{G}(z)}{1 + \Upsilon(z) + \alpha(z)\epsilon(z) - \bar{G}(z)}. \tag{6}$$

This differs from the standard Saez (2001) result by the term $\Upsilon(z)$. This new term is defined as $\Upsilon(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z)$, and consists of two parts. The magnitude of the inequality externality η is present in both. If η is large and positive, inequality is a significant negative externality (a public bad). If it is negative, inequality is a positive externality (a public good). The parameter $\kappa(z)$ denotes the weight of the individual at the tax bracket z in the inequality metric, and $\bar{\kappa}(z)$ denotes the average inequality metric weight of everyone above the tax bracket. In the absolute Gini, $\kappa(z) = 2F(z) - 1$ and $\overline{\kappa}(z) = F(z)$. We also use several of the standard parameters from the optimal taxation literature, both in $\Upsilon(z)$ and in the tax formula as a whole: the local Pareto parameter $\alpha(z) = \frac{zf(z)}{1-F(z)}$, the elasticity of earnings $\epsilon(z)$ (with respect to $1 - \tau(z)$), and the average social welfare weight above z, denoted by $\bar{G}(z)$.[26] We will now discuss the intuition behind Equation 6.

There are two key effects of a marginal tax raise on post-tax income, and thus two effects of a marginal tax raise on post-tax absolute income inequality.[27] These two effects correspond to the behavioral response and the mechanical effect from the classical OIT literature. The behavioral response captures how a small tax increase leads each agent located at that tax bracket to shift their work decision towards leisure. The mechanical effect captures how every agent above the tax bracket is taxed more without changing their work decision (or changing it in a limited fashion if there are income effects). These two channels both have effects on revenue and post-tax income inequality.

In the traditional literature, the key effect of each channel is that on tax revenue. This revenue is usually assumed to be redistributed equally to everyone; in general, more revenue implies more redistribution. The behavioral response represents a tax revenue loss, while the mechanical effect represents a tax revenue gain. The two terms together represents a revenue collection trade-off, the sufficient statistics of which can be empirically estimated. We will discuss these consequences as *revenue effects*.[28]

These two channels also impact post-tax income inequality directly, which is not considered welfare-relevant in traditional models. The mechanical effect always decreases inequality, as it redistributes income from those above a certain tax bracket equally to every agent. The behavioral effect, however, increases or decreases post-tax income inequality *depending on the location of*

[26] $\alpha(z) = \frac{zf(z)}{1-F(z)}$ is a distributional measure which becomes constant in a Pareto distribution. In the Rawlsian min-max framework, $\bar{G}(z) = 0$. See Saez (2001) for further discussion on these variables.

[27] Any income redistributed is given equally to all agents, which does not lead to any change in the absolute inequality metrics we use; thus we can focus on where post-tax income is reduced. We note that the upcoming intuition is largely the same with standard non-absolute inequality metrics. There would only be one minor difference; the behavioral channel would be somewhat less inequality-reducing due to the reduction in average income that follows from the behavioral responses.

[28] There is also a negative welfare effect of those who pay higher tax rates due to the mechanical effect, represented by $\bar{G}(z)$ in Equation 6, which we will ignore for now as it does not change the intuition below significantly.

the tax bracket. The uneven impact of the behavioral response is a key difference between the traditional revenue effects and the new equality impacts and why our novel policy implications are disproportionately localized at the top. The summary of this discussion is in Table II.

The two new terms introduced into the optimal tax formula through $\Upsilon(z)$ corresponds to these two effects. We will now discuss each in detail.

Table II

The Effects of a Small Tax Increase on Revenue R and Inequality $\bar{\theta}$

		Bottom incomes	Middle incomes	Top incomes
Behavioral response	Revenue effect[†]	$\longleftarrow$	Decreases R	$\longrightarrow$
	Inequality impact[†]	Increases $\bar{\theta}$	Small / no change to $\bar{\theta}$	Decreases $\bar{\theta}$
Mechanical effect	Revenue effect[‡]	$\longleftarrow$	Increases R	$\longrightarrow$
	Inequality impact[⌉]	$\longleftarrow$	Decreases $\bar{\theta}$	$\longrightarrow$

†: The behavioral response always decreases revenue, as individuals in the tax bracket shift away from work into leisure.

†: The behavioral response changes the work decision of the individuals in the tax bracket, which changes incomes. A tax raise on the bottom decreases the bottom agents' incomes, which increases inequality. A tax raise on the middle decreases the middle agents' incomes, with little to no inequality effect. A tax raise decreases the top agents' incomes, which decreases inequality.

‡: The mechanical effect always increases revenue, as individuals above the tax bracket have a higher average tax rate yet do not change their work decisions.

⌉: The mechanical effect always decreases inequality, as it redistributes a fixed amount of income from every individual above the bracket equally to every individual.

Note: The table describes the effect each channel exerts on inequality $\bar{\theta}$ and tax revenue R through a small marginal tax raise in the specified distributional location.

The behavioral response: A Pigouvian tax The first term, $\eta\alpha(z)\epsilon(z)\kappa(z)$, comes from the behavioral responses of the individuals who are located at income z. These agents work less due to the tax increase. The classical consequence of this response is that tax revenue is reduced, which is true no matter the location of the tax raise. In Equation 6 this is represented by $\alpha(z)\epsilon(z)$ in the denominator, which always reduces the optimal marginal tax rate.[29]

The equality impact, on the other hand, is conditional on the location of the individual. If agents at the bottom shift into leisure, their income decreases and income inequality increases. If agents in the middle shift do the same, there is little to no effect on income inequality. And for agents at the top, a shift into leisure *decreases* income inequality. Unlike in the traditional case, this implies a potentially positive welfare consequence of the behavioral responses in many tax brackets.

[29] The local Pareto parameter $\alpha(z) = \frac{zf(z)}{1-F(z)}$ can be understood as a measure of the relative strength of the mechanical effect and behavioral response. The numerator amplifies the behavioral channel and the denominator amplifies the mechanical channel. Although we include it in the behavioral response for mathematical simplicity, it affects both channels.

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 itself never negative, all else equal (upholding the Pareto principle). The optimal outcome is for individuals who make socially suboptimal labor choices to substitute into leisure, keeping their utility high.

The term corresponds to a Pigouvian tax designed to correct the individual's socially suboptimal labor decision. This suboptimality differs in magnitude and direction based on the position of the individual, and thus the optimal tax change from this term has different signs across the distribution. As an example, if we are examining an agent near the top in a negative inequality externality framework, their unbiased labor choice is skewed towards increasing individual income at a social cost. As $\kappa(z) > 0$ and $\eta > 0$, the optimal marginal tax rate on the agent is thus higher than in a no-externality framework; the new term makes the individual internalize part of the cost their high income places on society. Similarly, if we are in a positive externality framework such that $\eta < 0$, the agent will be subsidized to internalize the positive effect their increased income has on society.[30]

The term is affected by four parameters. First, how the agent affects inequality, represented by their weight in the inequality metric $\kappa(z)$. If the agent has a larger effect on the inequality metric, the optimal tax effect is likewise increased. Subsequently this term is large at the ends of the distribution (working in opposite directions at the top and bottom). Second, how inequality affects other agents, represented by the externality magnitude η. If other agents are significantly affected by inequality, the tax change will be larger. Third, the degree to which agents substitute away from work when taxed, represented by the elasticity $\epsilon(z)$. If agents substitute more to leisure, the equality impact of the tax increase is stronger. It follows that the term is largest when elasticities are *high*. Fourth, the total amount of agents at the tax bracket z, represented by the distributional term $\alpha(z)$. If there are more agents in the tax bracket, such that $\alpha(z)$ is large, there is a greater inequality impact and the optimal tax changes are larger.

These last two factors imply that the standard intuition from the revenue channel – where a high elasticity and a high $\alpha(z)$ leads to a low tax rate – is partially reversed in our framework. In particular we draw attention to the elasticity case. In the standard framework, high elasticities imply that 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.

This Pigouvian term invalidates three classic results from the literature based on Mirrlees (1971). These original results are fragile, and change with many small modifications to the model – see Stiglitz (1982) and Saez (2001) for examples. As such, these changes are not very surprising. Still, they are intuitively appealing, and as such we list them here:

1. Sadka (1976) and Seade (1977) observed that the marginal tax rate should be zero at the top of the income distribution. Known to be true only locally in standard models (Tuomala et al.,

[30] We note that this term exists specifically due to our choice of an *income* inequality externality. If the externality was in terms of utility, the behavioral response would not change the externality and the term would not exist.

1990; Saez, 2001), it is untrue even locally when adding an inequality externality. Reducing the income of the top-earner has become a social cost or benefit in itself, and should be a subsidy or tax depending on the direction of the inequality externality. The optimal marginal tax rate at the top of a bounded distribution is the following;

$$\tau(z) = \frac{\eta\kappa(z)}{1+\eta\kappa(z)}. \tag{7}$$

Which can be either a tax if inequality is a negative externality ($\eta > 0$) or a subsidy if inequality is a positive externality ($\eta < 0$).

2. Seade (1977) found that the marginal tax rate should be zero at the bottom of the income distribution, given that everybody works. The inequality externality negates this in a similar fashion.

3. Seade (1977) argued that the optimal marginal tax rate should be between zero and one. Given the above results this is no longer true – one can have negative rates both at the top and bottom.

These modifications to the classic OIT results are intuitively appealing. In particular, the change to the zero marginal tax rate at the top result is notable. We argue that this controversial result shows an intrinsic limitation of the Mirrlees (1971) model. In the classic model equality in itself is not valued by individuals, who are individually indifferent to distributional changes. As such, the income of the top agent is irrelevant for the rest of society unless it can directly contribute to tax revenue (and thus redistribution). This is contrary to intuition, and is particularly notable when considering extremely high incomes; whether in a positive or negative fashion, such incomes are likely to affect other agents. This is taken into account in our model, which imposes a tax or subsidy on the top agent depending on the direction of the externality.

The mechanical effect: An increased taste for (in)equality The second term, $\eta\bar{\kappa}(z)$, is from the mechanical effect on the agents located above income z. These agents have an unchanged marginal tax rate, so their work choice remains the same.[31] However, as their average tax rate increases, their post-tax income decreases. The classical consequence of this response is that tax revenue is increased, which is true no matter the location of the tax raise.

The equality impact comes from the post-tax income decrease of these agents. In sum, income collected from those above the marginal tax increase is redistributed equally to everyone. This decreases absolute inequality by definition almost no matter where the tax raise occurs. The sole exceptions are where no effective revenue is gathered; at the very top, where there are no agents above, and at the very bottom, where every agent is above. In every other case, the mechanical effect decreases inequality which leads to a welfare change in our framework.

How much this impacts optimal marginal tax rates depends on the average weight of the agents above the tax bracket in the inequality metric ($\bar{\kappa}(z)$) as well as how valuable or costly reductions to inequality are (η) and how many agents are above the bracket (which contributes to $\alpha(z)$).[32] Since

[31] This is due to the assumption of no income effects. The intuition remains similar with income effects.

[32] $\alpha(z)$ contributes to both channels. See footnote 29.

the inequality impact from the mechanical effect functions similarly to the associated revenue effect, the new term is similar to the old, represented by the numerical constants in the numerator and denominator. The standard mechanical effect term is dampened or amplified by a multiplicative factor dependent on how inequality changes, $\bar{\kappa}(z)$, and whether or not this is welfare-enhancing, η.

As the individuals' work decision is unaffected by the mechanical effect, this term indicates the increased social willingness to change inequality levels absent any other changes. It has the same sign as the inequality externality η, as $\bar{\kappa}(z)$ is always positive for all inequality metrics with monotonically increasing weights. For the Gini, $\bar{\kappa}(z) = F(z)$. Assuming a negative (positive) inequality externality, the full term unambiguously increases (decreases) the marginal rate in every tax bracket except at the very top and at the very bottom. The term exists whether or not the agent makes the socially optimal work decision, and can be approximated by appropriate social weights.

The externality thus introduces two new terms to the optimal tax formula. The first term is a result of agents shifting their labor income into leisure, internalizes the externality, and has an ambivalent sign depending on the tax bracket location. The second term is a result of the redistribution of income from agents above the tax bracket, symbolizes the increased social willingness to pay for (in)equality, and always has the same sign as the inequality externality.

Both new terms always change after-tax income inequality in the direction of the externality. If we define progressivity as a lower after-tax Gini coefficient (Piketty and Saez, 2007), the resulting optimal tax rates with a negative (positive) inequality externality are unambiguously more progressive (regressive) than the standard case. This shows an intuitive result; if inequality is considered a public bad, optimal income tax rates are more progressive than those previously found in the literature. If inequality is considered a public good, optimal income tax rates are more regressive than those previously found in the literature.

We can summarize the small perturbation method as follows. To find the social optimum, we equalize the welfare impacts of a small tax raise through the revenue channels dB and dM and the equality channels $d\bar{\theta}_M$ and $d\bar{\theta}_B$ (where we assume a Rawlsian SWF for simplicity[33]):

$$dM + dB + d\bar{\theta}_M + d\bar{\theta}_B = 0 \tag{8}$$

The revenue channels are simple; dM is positive and dB is negative. If inequality is a negative externality, the sign of $d\bar{\theta}_M$ is always positive. The sign of $d\bar{\theta}_B$ changes across the distribution. If we again assume a negative externality, it is negative near the bottom, near-zero around the middle, and positive near the top. This illustrates the trade-off between choosing equality levels and maximizing tax revenue.

In total, the sign change to the optimal top marginal tax rates from introducing the equality concerns is ambiguous at the bottom, where the equality impacts of the mechanical and behavioral channels are in opposition. At the top the change to the optimal tax rates is unambiguous, as the

[33] Introducing dW in the equation makes no significant difference; it is always negative and does not interact with the equality terms.

two channels work together; resulting rates are higher with a negative income inequality externality and lower with a positive income inequality externality. These are general results that do not depend on most of the assumptions we use for simplicity.[34]

III.B Numerical Simulations

In this section we use numerical calculations to find optimal marginal tax rates in the presence of an income inequality externality. The main focus of the numerical simulations will be on how the inequality externality changes the results from the no-externality case.[35] We use the analytic solution from Appendix C throughout, which is functionally identical to the method in the preceding section and avoids the problem of an endogenous pre-tax income schedule.[36] We assume quasi-linear utility, a constant labor elasticity, and a linear homogenous inequality externality.[37]

Method In the traditional optimal tax literature, tax rates are largely determined by three factors; (i) labor or earnings elasticities, (ii) the social welfare function, and (iii) the shape of the wage-earning ability distribution (see Mankiw et al. (2009)).

The first of these factors are the labor elasticities, which potentially vary over the distribution. These will be kept constant and homogenous for simplicity in our analysis. We assume that the elasticity of labor supply is constant at $E_L = 0.3$ for all income levels, a reasonable mid-range value from empirical estimates. This choice does not significantly impact the analysis.

The second factor is the social welfare function. To span the range of non-increasing social welfare functions we use two extremes; (i) the Rawlsian minmax, which implies that the objective function of the government is to optimize the welfare of the worst-off member of society, and (b) a fully Utilitarian function, implying that the utility of every agent is equal.

The third factor is the shape of the wage-earning ability distribution $F(n)$. Our main specification uses empirical survey data for the 2018 U.S. wage distribution gathered from the Annual Social and Economic Supplement of the Current Population Survey.[38] The underlying density distribution $F(n)$ was extracted using Kernel density estimation. Because survey data is incomplete towards the top, we also assume that the wage distribution approximates a Pareto distribution for wages above \$320/hour with a constant parameter estimated from the top. This is $\alpha(n) = 1.9$,

[34] Under the assumptions we use to find Equation 6 we can also note that optimal marginal tax rates above the median wage always increase (decrease) as compared to the standard case given a negative (positive) inequality externality.

[35] See the discussion in Saez (2001), among others, for a numerical exploration of the standard parameters.

[36] This does not significantly affect the results. Our no-externality results are nearly identical to those found in Saez (2001) and others.

[37] This implies a monotonic transformation of the following utility function:

$$U(x, h, \bar{\theta}) = log\left(x - \frac{h^{\left(1+\frac{1}{E_c}\right)}}{\left(1+\frac{1}{E_c}\right)} - \eta\bar{\theta}\right) \tag{9}$$

[38] Microdata were collected with IPUMS (Flood et al., 2018). Total wage income was divided by the average hours worked in a year to find the hourly wage distribution for individuals aged between 21 and 66 years. Individuals with no or negative wage income were excluded.

Figure I: Weights of Inequality Metrics

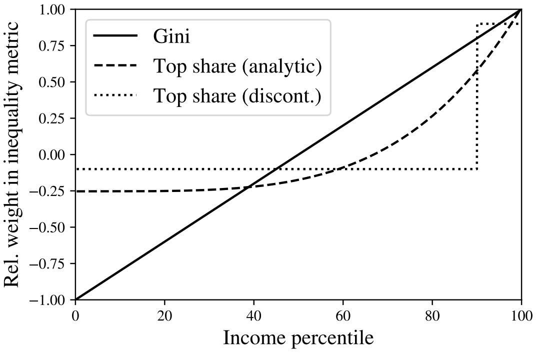
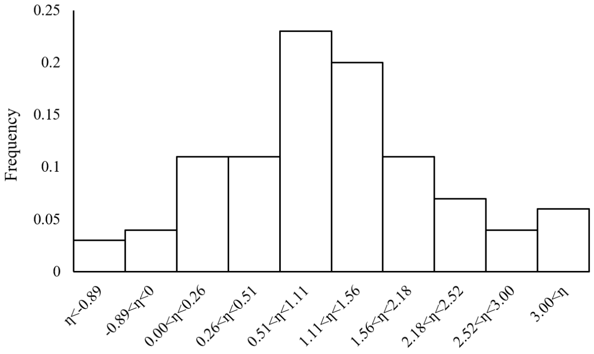


Figure II: Estimated η_G (Carlsson et al., 2005)



Notes: Figure I 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 Figures III and IV, respectively). Figure II 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).

very close to the $\alpha(n) = 2.0$ used in Saez (2001). Slightly more than 0.5% of all income-earners are affected. In addition to this empirical wage-earning ability distribution, we also present two standard theoretical distributions in Appendix D.

These choices decide the shape of the no-externality optimal tax function. The externality necessitates two additional choices; the inequality metric and the size and direction of the externality.

The two inequality metrics we use were introduced in Equations 3–5 (with $q = 4$ in the latter case). Their weights $\kappa(z)$ in the general expression $\bar{\theta} = \int_{\underline{z}}^{\bar{z}} \kappa(z)x(z)dF(z)$ are plotted in Figure I. The figure shows the relative weight of the income of any agent when calculating the specified inequality metric. It also shows the weights used in the top 10% income share for comparison, which is discontinuous and thus not usable in an analytical setting. We use the Gini coefficient in the main specification and the top income share-based metric in the main robustness check. Other inequality metrics are examined in Appendix D.II.

Given the inequality metric we need to choose values for the inequality externality magnitude. As there are unavoidable empirical challenges in calibrating such a number, we do not aim to strongly argue for any one parameter value. We instead use a range of realistic values to illustrate the potential tax policy consequences of various income inequality externalities.

To find such a range of η we present estimates based on data from Carlsson et al. (2005). The work uses a survey design to find macroeconomic inequality aversion estimates in Swedish university students. 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.

The values of η depend on which inequality metric is chosen to be relevant for the externality. We denote the values calculated for the Gini coefficient as η_G. The median respondent in the survey has approximately $\eta_G = 1.00$. A majority of respondents have $0.26 < \eta_G < 2.18$.[39] The full distribution is presented in Figure II.[40] 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.

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?*[41]

Answering the question creates equivalent incomes for differing inequality levels. These equivalent incomes for Denmark and the United States, and their corresponding η when using the Gini, are shown in Table III. 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. If $\eta_G = 0$, the agent is indifferent without any change to their income. The change in income compensates for any inherent dislike of inequality as well as any potential inequality effects, i.e. any macroeconomic or societal changes that are caused by the change in inequality.

Table III

The Magnitude of Inequality Externalities η_G

	$\eta = -0.5$	$\eta = 0.0$	$\eta = 0.5$	$\eta = 1.0$	$\eta = 2.0$	$\eta = 3.0$
U.S. Income Multiplier	0.94	1.00	1.06	1.13	1.25	1.38

Note: 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 Equation 9.

Based on these two 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.

We check that the individual's second-order conditions hold in every simulation using two different methods; first we ensure that earnings increases over ability (Lollivier and Rochet, 1983), and second we numerically ensure that the incentive compatibility constraint is satisfied for every agent.

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

[40] Using inequality aversion instead of a direct externality estimate means that we are using for preferences to proxy for effects – see the discussion in Section II. 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 homogeneity assumption matches the respondents' perceived utility functions. All these reasons contribute to why we are using a *range* of η.

[41] 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.

Main Results: The Gini Externalities Our main specifications, using the Gini as the post-tax income inequality metric, are presented in Figures III. The introduction of even a small income inequality externality substantially changes the optimal tax structure. The effect is larger towards the top of the income distribution.

In the Rawlsian case, the top marginal tax rate increases from 70% to 90% when assuming a moderately large negative inequality externality, $\eta_G = 2.0$ (see Table III). 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 85%.

With a small positive inequality externality ($\eta_G = -0.5$), the optimal top marginal tax rate is only 40% and approaches zero around the 97th percentile. The effects of the positive inequality externality are almost entirely located above the 90th percentile. All simulations have a small decrease in optimal tax rates around the 90th-95th percentiles; this is due to the well-known decrease of the local Pareto parameter of the empirical wage distribution around these values.

In the Utilitarian case, the very top marginal tax rates converge to the Rawlsian rates. Below the top, the resulting optimal marginal tax rates are lower than the Rawlsian case. In particular, a positive externality of $\eta_G = -0.5$ leads to negative optimal marginal tax rates for income earners between the 84th and 98th percentiles.

At the bottom of the distribution, all Rawlsian optimal rates converge to very high values. This is 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 across the distribution.

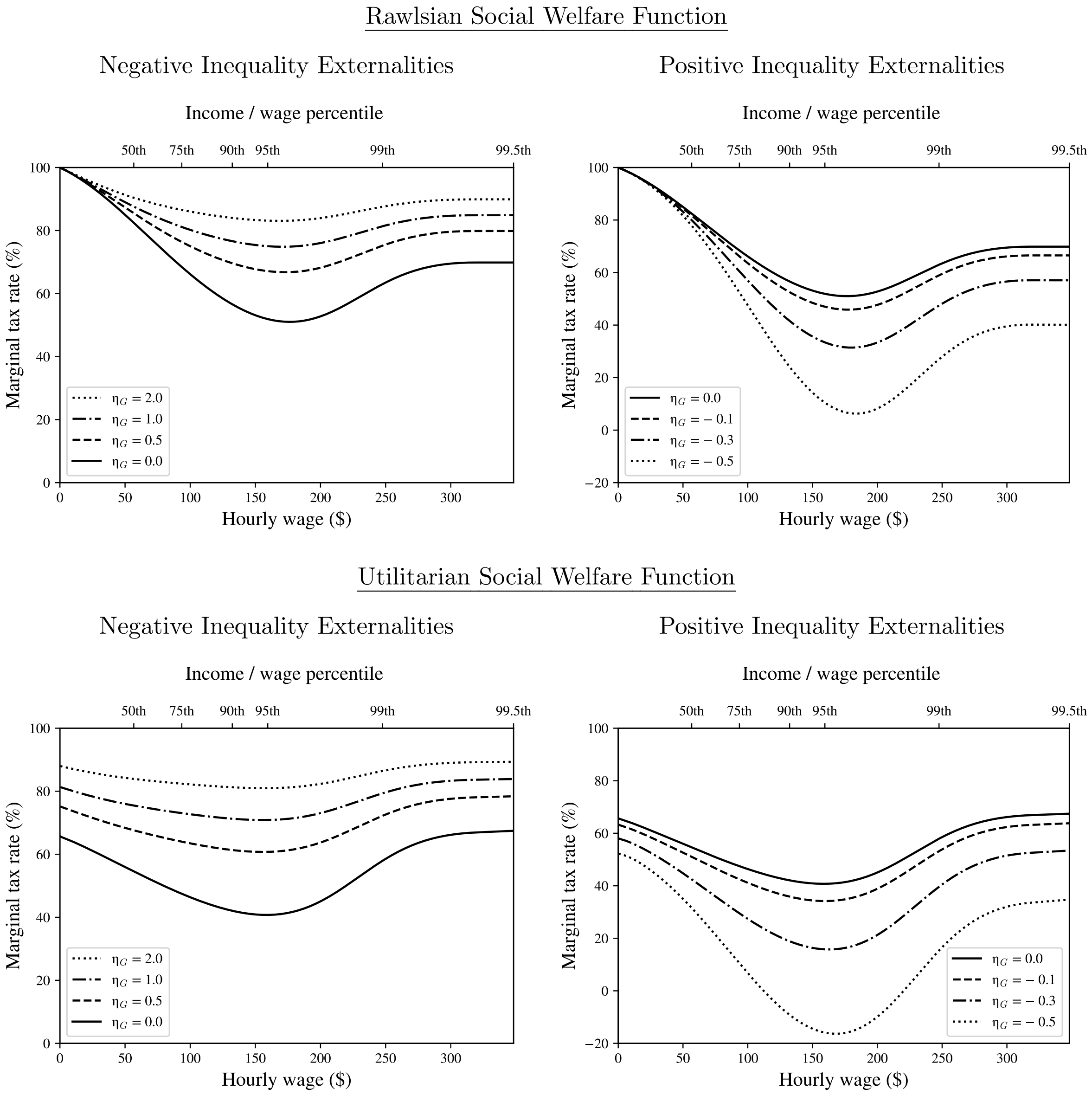
In the Utilitarian case, the utility losses of those above discounts the benefit of the tax raise. Very high bottom marginal tax rates are thus less appealing. Since the cost-benefit trade-off is less clear, the effects of the inequality externality are also more visible. The distributional equality effect of the mechanical effect is still large, even though it is utility-discounted when considering social welfare. As such, between the two introduced equality effects, the mechanical effect dominates the behavioral response at the bottom under our parameter choices. This results in the optimal marginal Utilitarian rates being increased (decreased) across the distribution with a negative (positive) externality.[42] This result is not universal, and the effect of the externality at the bottom is usually smaller than in this Utilitarian case due to the counteracting behavioral response.[43]

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

[42] The behavioral response is not visible in the graphs, but it dampens the increase or decrease in tax rates.

[43] The Utilitarian case with no income effects has among the least top-heavy distributional 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 D.I, a negative externality *decreases* optimal marginal tax rates at the bottom.

Figure III: Optimal Marginal Income Tax Schedules with Gini Inequality Externalities



Rawlsian Social Welfare Function

Negative Inequality Externalities

Positive Inequality Externalities

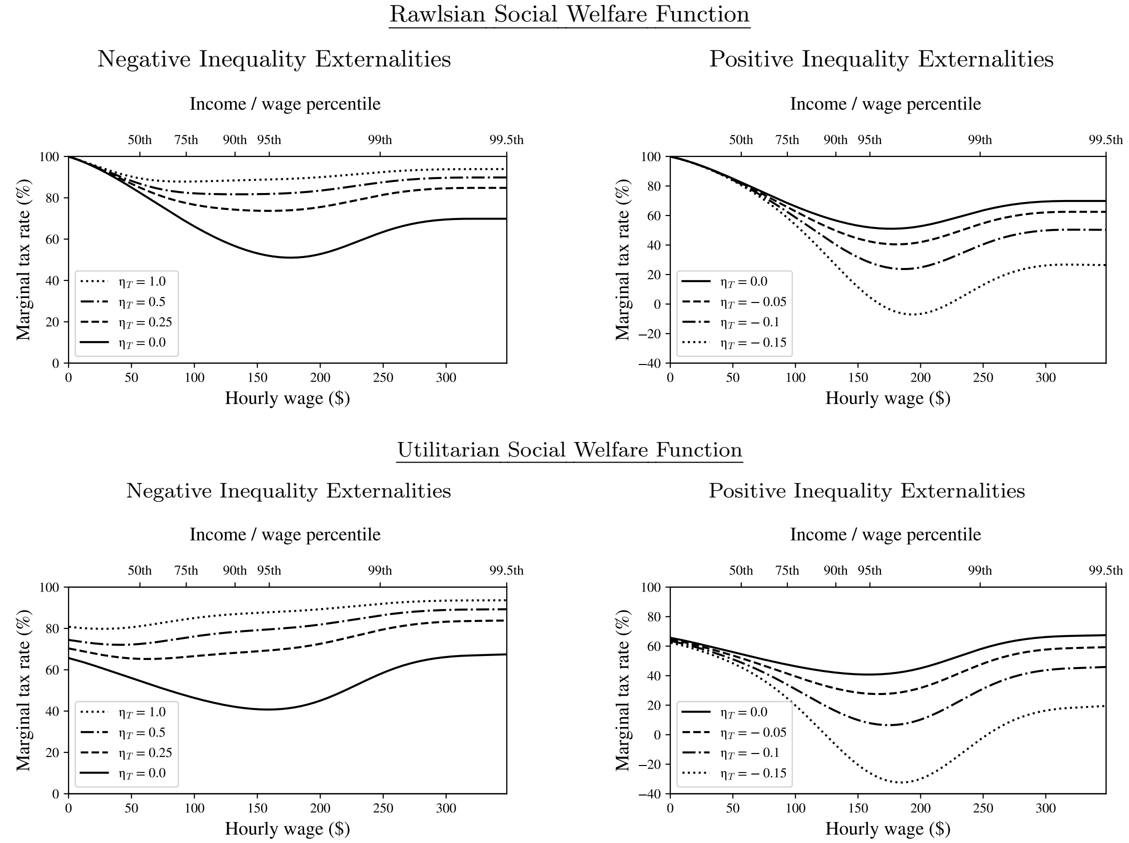
Utilitarian Social Welfare Function

Negative Inequality Externalities

Positive 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 Rawlsian (above) and Utilitarian (below). The two cases converge when moving towards the top. 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 III. Note the different scales of the vertical axes between the negative and positive externalities.

Figure IV: Optimal Marginal Income Tax Schedules with Top Share Inequality Externalities


Rawlsian Social Welfare Function

Negative Inequality Externalities

Positive Inequality Externalities

Utilitarian Social Welfare Function

Negative Inequality Externalities

Positive Inequality Externalities

Notes: Optimal marginal tax rates for various top share-based inequality externalities with magnitudes η_T where inequality is either a negative externality (left) or a positive externality (right). The social planner is Rawlsian (above) and Utilitarian (below). The two cases converge when moving towards the top. Empirical estimates indicate $\eta_T = 0.5$. The solid line, $\eta = 0$, is the standard no-externality case. Note the different scales of the vertical axes between the negative and positive externalities.

Robustness: Top Income Share Externalities The choice of the inequality metric naturally influences our results. And 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 in Figure IV by using the top income share metric shown in Figure VII as the relevant inequality measurement. This inequality metric is defined as $q = 4$ in the general top income share metric family $\kappa(z) = (q+1)F(n)^q - 1$, $q \in \mathbb{N}$.

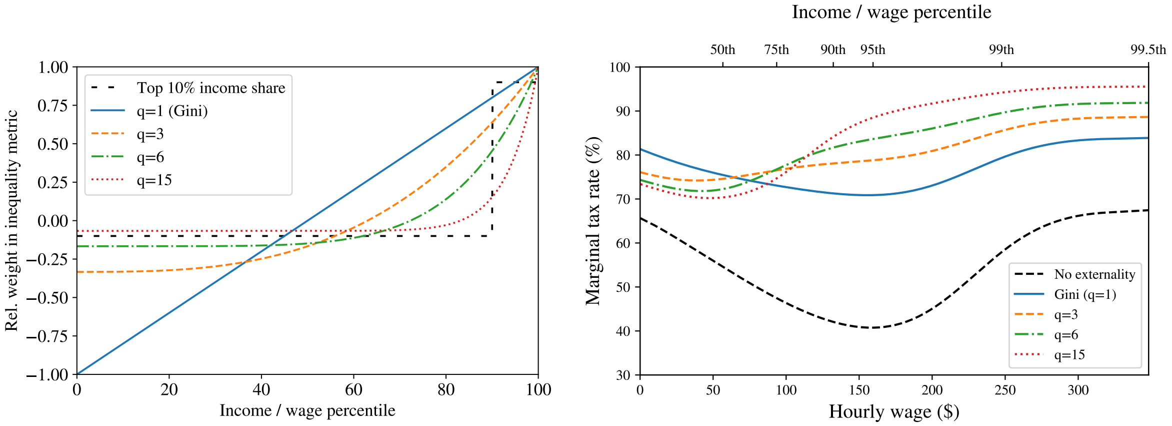
The externality effects are larger at the top and smaller at the bottom when using the top income share metric. In the Rawlsian case, the optimal top marginal income tax rate goes from 70% in the no-externality case to 94% when $\eta_T = 1.0$ (comparable to $\eta_G = 2.0$ in Figure III). For $\eta_T = 0.50$, the value closest to the empirical externality estimate taken from Carlsson et al. (2005), the optimal top marginal tax rate is 90%. Further, if inequality is a positive externality, top tax rates are often negative around the top, even in the Rawlsian case. If $\eta_T = -0.15$ in the Rawlsian case, optimal marginal tax rates begin at near a hundred percent and go below zero between the 96$^{\text{th}}$ and the 99$^{\text{th}}$ percentiles – the optimal top marginal tax rate is 26%. At the bottom of the distribution, optimal marginal rates again approach one hundred percent.

In the Utilitarian case, the effect of the externality on optimal top rates is also larger than when using the Gini (as expected, since the very top rates again converge to the Rawlsian case). In addition, the effects close to the top are larger and the effects close to the bottom are smaller. If $\eta_T = -0.15$ in the Utilitarian case, the optimal marginal tax rate reaches -32% near the 97th percentile and is negative between the 87th and 99th percentiles. Near the bottom, the effects are relatively small. The negative externalities increase optimal marginal tax rates by around fifteen percentage points at most, whereas the positive externalities barely impact the optimal bottom marginal tax rates.

That the effects of the externality are increasingly concentrated towards the top of the distribution with increasing q, i.e. when we move away from the Gini towards a top income share, is a general result. We show this more clearly in Figure V. There we plot more inequality metrics from the top income share family alongside their resulting optimal tax rates. The externality is kept constant at the upper bound of the median inequality aversion range from Carlsson et al. (2005). Figure IV used $q = 4$; here we show the effect of moving from $q = 1$ (the Gini) to values increasingly focused on top incomes (up to $q = 15$). The no-externality case is shown as a reference. The ability distribution is the empirical wage distribution, and the SWF is Utilitarian.

Figure V

Varying the Inequality Metric with a Fixed Externality Magnitude



Note: 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 social planner is Utilitarian. The productivity distribution is the empirical wage distribution. The black dotted line is the standard case of no inequality externality. The elasticity of labor E_L is 0.3.

The optimal tax changes are larger near the top with increasing q, which should not be surprising given the increasing weight of top incomes in the inequality metric. It is also noticeable, however, that the effects near the bottom are reduced. This is due to the mechanical effect being less powerful at the bottom; redistributing from everyone above is less impactful for inequality-reduction if everyone in the lower half are weighted somewhat equally. Overall, using top income shares further concentrates the effect of the externality towards the top.

With other inequality metrics, such as those in the S-Gini family, results are overall similar. This is further discussed in Appendix D.II. 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.

III.C Equality concerns: Top tax rates

Equality concerns – the consequence of the inequality externality – come in addition to the revenue concerns usually discussed in the OIT literature. Their policy importance differs based on income bracket. In particular, as we have discussed in the preceding sections, equality concerns have a large effect on the optimal top tax rate.

Equality concerns and revenue considerations differ in their impacts across tax brackets. 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). In other social welfare functions the welfare benefit from revenue is usually relatively stable in the top half of the distribution. 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.

It follows that 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 tax rates – *if* they disagree on how inequality changes society. In keeping with the logic of inequality effects, this can be true even in the absence of jealousy and envy. Our numerical simulations in Section III.B strengthen this point.

Below we discuss specific findings related to these large impacts on optimal top income tax rates. First we show two real-world implications of our model, justifying observed policy arguments that cannot be rationally explained under standard revenue considerations. Second we discuss the existence of optimal rates higher than the revenue-maximizing Laffer rate.

1 Large variation in top rates: A maximum income, or the Rawlsian Conservative? OIT models are generally considered more accurate towards the top of the distribution. Top marginal income tax rates often converge to around $60-70\%$, even in the Rawlsian case. Although these numbers depend heavily on parameter specifications, heterodox assumptions are required for optimal rates below 50% or above 80%.[44]

As we have shown in the preceding sections, varying the value of the inequality-sensitivity parameter η has a large effect on the top optimal income tax rates. This variation is large even

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

when compared to the variation induced by changing standard parameter values. We examine this in Tables IV and V. These tables show how the optimal top tax rate varies with (1) combinations of Gini income inequality externality magnitudes η_G and the inverse local Pareto parameter $1/\alpha$ (Table IV) or (2) combinations of Gini income inequality externality magnitudes η_G and labor elasticity values E_L (Table V). The inequality externality induces changes that are generally larger than the effects from changing $1/\alpha$ and E_L. By changing η within reasonable bounds, the same Rawlsian social planner can find optimal top tax rates from near-zero to near-one. In other words, almost any top tax rate can be optimal depending on the magnitude of the inequality externality.

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, here represented by $\eta = 3.0$, 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 $\eta = -0.5$, 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 have been proposed in political discourse. In 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.

2 The Laffer Curve The central idea of the Laffer curve is simple and true; above a certain tax threshold revenue drops with increased taxation. However, the Laffer curve is often also described as an upper bound on sensible taxation. Laffer (2004) describes this as the "prohibitive range" of taxation, and Manning et al. (2015) argue that "one would not want a rate higher than the Laffer rate".

In the presence of an inequality externality the above statements could be either misleading or false. The externality negligibly changes agent behavior when there is a large number of agents, so the revenue-maximizing rate does not change. However, the welfare maximizing rate can change, and is in fact often above the Laffer rate given the public benefit of distributional changes.

As an example, consider a society with ten agents, one vastly more wealthy than the other nine. Given the desirability of equality, the welfare-maximizing top marginal rate can be higher than the revenue-maximizing rate, which is zero at the top according to standard results. The Rawlsian numerical simulations in Section III.B provides another example.

The optimal income tax rate can be higher than the revenue-maximizing rate both at the top (given a negative externality), and at the bottom (given a positive externality). Specifically, the optimal marginal income tax rate is higher than the revenue-maximizing marginal income tax rate

Table IV

Optimal Top Tax Rates, Inequality Externalities and Distribution Parameters

		Inverse top Pareto parameter $1/\alpha$											
		0.25	0.27	0.29	0.31	0.33	0.36	0.40	0.44	0.50	0.57	0.67	0.80
	-0.50	4	7	11	14	18	22	27	32	37	42	49	55
	-0.25	36	38	40	43	45	48	51	54	58	62	66	70
	0.00	**52**	**54**	**55**	**57**	**59**	**61**	**63**	**66**	**68**	**71**	**74**	**78**
	0.25	62	63	64	66	67	69	71	73	75	77	79	82
Sensitivity to inequality η	0.50	68	69	70	71	73	74	76	77	79	81	83	85
	0.75	73	73	74	76	77	78	79	80	82	84	85	87
	1.00	76	77	78	79	80	81	82	83	84	86	87	89
	1.25	79	79	80	81	82	83	84	85	86	87	89	90
	1.50	81	81	82	83	84	84	85	86	87	88	90	91
	1.75	83	83	84	84	85	86	87	88	89	90	91	92
	2.00	84	85	85	86	86	87	88	89	89	90	91	93
	2.25	85	86	86	87	87	88	89	89	90	91	92	93
	2.50	86	87	87	88	88	89	90	90	91	92	93	94
	2.75	87	88	88	89	89	90	90	91	92	92	93	94
	3.00	88	88	89	89	90	90	91	91	92	93	94	94

Note: Top marginal tax rates from Equation 6 with varying values of an inequality externality and the inverse local Pareto parameter $1/\alpha$ at the top. The social planner is Rawlsian. The elasticity of labor E_L is 0.3. The inverse local Pareto parameter $1/\alpha$ is approximately 0.5 at the top in empirical data (and in the remainder of the paper). The standard no-externality case is in bold.

Table V

Optimal Top Tax Rates, Inequality Externalities and Labor Elasticities

		Elasticity of labor E_L									
		1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	-0.50	0	3	6	10	14	20	27	37	50	69
	-0.25	33	35	37	40	43	47	52	58	67	79
	0.00	**50**	**51**	**53**	**55**	**57**	**60**	**64**	**68**	**75**	**85**
	0.25	60	61	62	64	66	68	71	75	80	88
Sensitivity to inequality η	0.50	67	68	69	70	71	73	76	79	83	90
	0.75	71	72	73	74	76	77	79	82	86	91
	1.00	75	76	76	77	79	80	82	84	88	92
	1.25	78	78	79	80	81	82	84	86	89	93
	1.50	80	81	81	82	83	84	85	87	90	94
	1.75	82	82	83	84	84	85	87	89	91	94
	2.00	83	84	84	85	86	87	88	89	92	95
	2.25	85	85	86	86	87	88	89	90	92	95
	2.50	86	86	87	87	88	89	90	91	93	96
	2.75	87	87	87	88	89	89	90	92	93	96
	3.00	88	88	88	89	89	90	91	92	94	96

Note: Top marginal tax rates from Equation 6 with varying values of an inequality externality and elasticity of labor E_L. The social planner is Rawlsian. The inverse local Pareto parameter $1/\alpha$ is 0.5 in these calculations. The elasticity of labor E_L is 0.3 in the remainder of the paper. The standard no-externality case is in bold.

if, using the framework in Equation 6,[45]

$$\eta\alpha(z)\epsilon(z)\kappa(z)+\eta\bar{\kappa}(z)>\bar{G}(z),$$

that is, if the equality effects of taxation are larger than the welfare effects. If $\eta=0$ the inequality externality does not exist and the statement never holds unless social weights are negative, the standard result. As $\kappa(n)$ goes from negative to positive with higher incomes, and η changes sign depending on the direction of the externality, it can hold either at the bottom (with a positive externality, $\eta<0$) or at the top (with a negative externality, $\eta>0$).

In the Rawlsian case, the right-hand side of Equation 10 is zero above the very bottom earner. Thus, using the Gini values, the inequality simplifies to

$$\frac{F(z)}{\alpha(z)\epsilon(z)}>1-2F(z), \tag{11}$$

which is independent of η and holds for any income above the median.[46]

The Mirrlees literature occasionally uses the revenue-maximizing rate as a necessary 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.

IV Conclusion

This paper has introduced the concept of an *inequality externality* and has particularly focused on an *income* inequality externality.

Most standard models of welfarist policy design implicitly assume that income inequality has no societal effects. As we have shown with microfounded examples, such effects likely exist and could be both numerous and impactful. They are often independent from individuals' personal feelings; if inequality increases crime, for example, even a selfish individual would prefer equality in the absence of other changes. Including such effects into simple welfarist models with only a combination of diminishing marginal utilities of income and social welfare weights is not possible. The inequality externality is thus intended as a simple and generalizable way to model these side-effects of economic inequality without having to specify the potentially numerous causal channels independently. The inequality externality concept itself is tractable and does not assume a direction to the externality, can include other-regarding preferences but does not require them, and can easily be extended to other dimensions such as wealth inequality or heterogeneous utility functions.

[45] In the most general framework, see Appendix C, this is equal to,

$$\gamma\left[\kappa(n)+\frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty\left[\frac{\kappa(p)}{u_{x(p)}}\right]f(p)dp\right]>\frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty\left[W'(U(p))\right]f(p)dp, \tag{10}$$

which represents the same intuition; the equality effects of taxation must be larger than the welfare effects.

[46] This is intuitive; the Rawlsian rate is the revenue-maximizing rate, and the incentive for equality increases tax rates at least above the median agent.

Introducing an inequality externality to the welfarist framework leads *equality or inequality itself* to become a policy goal. Individual labor decisions become socially suboptimal, and the marginal social welfare of individual income can become negative. Frameworks known for only being self-selection problems – including the optimal taxation problem – take on externality dimensions.

In the Mirrlees (1971) optimal income taxation model, the optimal non-linear tax structure becomes unambiguously more progressive with the introduction of a negative inequality externality. The two new terms in the optimal taxation formula correspond to the well-known mechanical effect and behavioral responses respectively, and represent (i) society's increased willingness to pay for redistribution, and (ii) the internalization of the individual externality on income.

We present three new insights to the optimal income taxation literature, all of which are relevant for real-world tax design.

First: Optimal top marginal tax rates are largely determined by the magnitude of the inequality externality. 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 (<30%) when inequality is a social good. Our median estimate is an 85% optimal top marginal tax rate. We thus find theoretical support for several policy arguments previously unsupported by economic theory, including a near-maximum income (with a large negative externality) or low top tax rates under a Rawlsian social planner (with a large positive externality). The findings also imply that different beliefs about the magnitude of the inequality externality could be a potential source of political disagreement. An intuitive explanation of this finding is that individuals at the ends of the distribution naturally impact inequality the most, but only those at the top can be specifically targeted by marginal tax rate changes. Mathematically, it follows from the top of the distribution being the only place where the two consequences of a tax raise, the mechanical effect and the behavioral responses, both affect inequality in the same direction (a reduction). The direction of the optimal tax rate change at the top is thus unambiguous given the direction of the inequality externality. At the bottom, where the two channels work have opposite equality impacts, it depends on model specification.

Second: The externality creates a trade-off between income inequality levels and tax revenue, which implies that the equality dimension of the optimal tax problem is more complex than simply choosing an appropriate social welfare function. The social planner must, at times, balance the benefit of higher tax revenue against the equality-related benefits from individuals shifting into leisure (from socially suboptimal high labor choices). This trade-off is particularly noticeable when inequality can change substantially with minimal revenue losses. In sum, even a Rawlsian social welfare function can put too low of a value on equality. Given that policy makers believe that inequality itself is concerning, the analysis presented here recommends more progressive taxes than those previously suggested by Saez (2001), Piketty et al. (2014), and others.

Third: The theoretical implications of the model change substantially, including welfare *benefits* from high income elasticities, as well as optimal rates above the revenue-maximizing Laffer rate. The former is driven by the welfare gains from individuals shifting away from suboptimally high labor choices. The latter is driven by individuals being willing to pay for lower (or higher) inequality.

Moreover, several results from the original Mirrlees (1971) model no longer hold.

Finally, given that many economic models are reliant on the assumption that inequality has no societal effects, the magnitude of our results could have widespread implications. We suspect that most of welfare-based economic theory has been lacking clear justication for this choice, and that further research on the topic will be fruitful.

Appendix

A Varying welfare weights

Another approach to introducing a dislike of inequality, common in the optimal income taxation literature, is varying the social welfare weights. The weights vary with utility such that the derivative of the SWF, $W'(U(n))$, is non-constant. The intuitive implication is that the welfare of the wealthy is weighted less. It is often presented as social inequality aversion, as it implies that the social planner values equality in itself.

There are three significant differences between this approach and the individual inequality externality we use in this paper.

First: Using social weights, the high income of one agent has no negative implications on others. In other words, the specification implies that there are no social externalities from agents' high income. A reduction of inequality is not beneficial *per se*; it is only beneficial if income is actually redistributed. This changes the implications of the exercise dramatically, from a pure self-selection problem (the standard problem) to an externality and self-selection problem (our problem).

Second: Using only social weights and absent other distortions, there is no difference between the optimality of the private and social labor supply choice. *Utility* is discounted, not *income*. Agents make the socially correct work decision.

Third: The social weights model imply that the social planner values utility unequally. Large parts of economic theory is based on the idea of a Utilitarian or Rawlsian social planner; moving to an inequality externality allows us to return to these assumptions while still allowing for inherent effects of inequality.

As we describe in the main section, there are three distinct ways to model the consequences of inequality. The cumulative effect of diminishing marginal utility, generalized social weights, and an inequality externality. These are distinct, occur through different mechanisms, and have different policy implications.

We now present a simple example to illustrate how an inequality externality can add nuance that cannot be found when only using social weights and the diminishing marginal utility of income. Imagine a world where one agent has seized the vast majority of income and uses this inequality of income to enjoy disproportionate (and socially damaging) political power. All other agents are equally poor. Now, imagine reducing the income of the oppressive ruler slightly, all else equal. We evaluate this change in the presence of *only* (i) risk aversion (diminishing marginal utility), (ii) a weighted social welfare function with non-negative weights, and (iii) an inequality externality.[47]

(i) Social welfare is unambiguously reduced, as the top individual's income decreases.

(ii) Social welfare is either reduced or kept constant – the top individual's income decreases, but they might have a zero social weight.

[47] The 'standard' case here is no risk aversion, a utilitarian welfare function, and no externality. For example, the first case will consider reducing the income of the top earner in a model with risk aversion, a utilitarian social welfare function and no externality.

(iii) The effect on social welfare is ambiguous. On one hand, the income of the top individual is reduced, reducing their utility and thus social welfare (if their weight is non-zero). On the other, income inequality is reduced, increasing every other agent's utility. The total effect on social welfare depends on the size of the inequality externality. In extreme cases, such as in this example, overall social welfare might *increase*.

More generally, diminishing marginal utility of income and social welfare weights present no intrinsic externality issues. As such, concentrated income gains lead to unambiguously non-negative welfare changes in standard models. Considering the current academic and social focus on inequality, this could be a troubling feature.

We note that the social weights discussed in this paper are in terms of utility. In certain situations income-based social weights can approximate the inequality externality, which illustrates a benefit of the generalizable social weights discussed in Saez and Stantcheva (2016). One example is when the utility function is separable in the externality term and there is a continuum of agents (as in Section III). In many other cases, however, an inequality externality would lead to other issues that could not be modeled by such weights alone (most notably changes to agents' behavior).

We present a proof below to show that appropriate utility-based social welfare weights cannot supplant an inequality externality.

A.I. Proof: The inequality externality cannot be approximated by social weights

The social planner aims to maximize:

$$W = \int_i g_i U(x_i, h_i, \theta(\boldsymbol{x}))di$$

Assume that g_i can have variation (social weights), and that $\frac{\partial U}{\partial \theta} \neq 0$ and $\frac{\partial \theta(\boldsymbol{x})}{\partial x_i} \neq 0$ (an inequality externality exists). x_i is income, h_i is hours worked, and $\theta(\boldsymbol{x})$ is inequality as a function of all incomes $\boldsymbol{x}$.

It follows from the social planner's first-order conditions for x_i and h_i that for all $g_i \neq 0$:

$$\frac{\partial U(x_i, h_i, \theta(\boldsymbol{x}))}{\partial h_i} = \frac{\partial U(x_i, h_i, \theta(\boldsymbol{x}))}{\partial x_i} + \frac{1}{g_i}\int_j g_j \frac{\partial U(x_j, h_j, \theta(\boldsymbol{x}))}{\partial \theta(\boldsymbol{x})}\frac{\partial \theta(\boldsymbol{x})}{\partial x_i}dj \tag{12}$$

We proceed with a proof by contradiction. Say we want to approximate the effect of the inequality externality with new social weights $\hat{g}_i$ without explicitly including θ in the utility function, otherwise keeping the utility function the same. Denote this new utility function $\hat{U}$. If so, $\frac{\partial \hat{U}(x_j,h_j)}{\partial \theta(\boldsymbol{x})} = 0$ and the second term on the right-hand side of Equation 12 is zero. The solution to the social planner's problem would thus involve $\frac{\partial \hat{U}(x_i,h_i)}{\partial x_i} = \frac{\partial \hat{U}(x_i,h_i)}{\partial h_i} \,\forall\, \hat{g}_i \neq 0$, which is equivalent to $\frac{\partial U(x_i,h_i,\theta(\boldsymbol{x}))}{\partial x_i} = \frac{\partial U(x_i,h_i,\theta(\boldsymbol{x}))}{\partial h_i} \,\forall\, \hat{g}_i \neq 0$. However, in the correct solution we are trying to approximate, $\frac{\partial U(x_i,h_i,\theta(\boldsymbol{x}))}{\partial x_i} \neq \frac{\partial U(x_i,h_i,\theta(\boldsymbol{x}))}{\partial h_i} \,\forall\, g_i \neq 0$. This implies that $g_i \neq 0 \rightarrow \hat{g}_i = 0$, which cannot be the case. Thus there is a contradiction. This follows from the externality creating a difference between the optimal individual and social work decisions, which cannot be introduced through discounting utility with social weights.

An extension shows that the externality cannot be approximated by the individual parameters in the utility function. If x_j is changed, Equation 12 implies that it will impact the FOC for i. In the modified solution with $\hat{U}$, it has no effect. To correctly specify $\hat{U}(x_i, h_i)$, one would need x_j or h_j. This would amount to including a distributional parameter $\theta(\boldsymbol{x})$ in the individual utility function, again a contradiction.

B Small Perturbation Solution to the OIT Problem

The core part of this approach follows Saez (2001) and Saez and Stantcheva (2016).

We introduce a small tax reform $d\tau_z$ where the marginal income tax is increased by $d\tau$ in a small band from z to $z + dz$. The reform mechanically increases average tax rates on everyone above this band. This is the mechanical effect of taxation, and collects $dz\partial\tau$ from $1 - F(z)$ agents above z under the assumption of no income effects. Thus it collects $[1 - F(z)]\, dz\partial\tau$ revenue. For each $dz\partial\tau$ collected, however, inequality also changes. The magnitude of this change per agent above differs based on which agent is considered. Noting that income rank $\kappa(z)$ does not change, each decrease in one unit of post-tax income at z changes absolute post-tax income inequality by $\kappa(z)f(z)$ (from Equation 3).[48] The mechanical effect thus has a differing equality effect of $\kappa(z_j)f(z_j)dz\partial\tau$ at each point j above z, where z_j is the income of the agent and $f(z_j)$ is the number of agents at this point, and $\kappa(z_j)$ is that agent's weight in the inequality metric. As the income change of each agent above z is equal, we can define the average inequality weight above as $\overline{\kappa}(z)\,[1 - F(z)] = \int_{\{j:z_j>z\}} \kappa(z)f(z)dj$ and write that the mechanical effect changes income inequality by $d\bar{\theta}_M = -\overline{\kappa}(z)\,[1 - F(z)]\, dz\partial\tau$.[49]

Those who are located in the small band between z to $z+dz$ have a behavioral response to the tax change. They work less, and reduce their pre-tax earnings by an amount $\partial z = -\epsilon(z)z\partial\tau \big/ (1 - \tau(z))$. $\epsilon(z)$ is the elasticity of earnings z with respect to $1 - \tau(z)$. There are $f(z)dz$ individuals in the tax bracket who were taxed at $\tau(z)$ before the perturbation, so total revenue decreases by $-dz\partial\tau \cdot \epsilon(z)zf(z)\tau(z) \big/ (1 - \tau(z))$. This change in total earnings is moderated by an effect $(1-\tau)/\tau$ for the inequality effect, as we are interested in the post-tax income decrease and not the tax revenue decrease.[50] Additionally we must multiply by the agents' weight in the inequality metric $\kappa(z)$. The behavioral response thus has an effect on the post-tax income inequality metric as $d\bar{\theta}_B = -\kappa(z) \cdot dz\partial\tau \cdot \epsilon(z)zf(z)$.

The total revenue effects are:

$$dR = dz\partial\tau \left(1 - F(z) - \epsilon(z)zf(z)\tau(z) \big/ (1 - \tau(z))\right)$$

The direct welfare effect through the individual income channels is $\int_j g_j dR dj$ for $z_j \leq z$ and $-\int_j g_j(\partial\tau dz - dR)dj$ for $z_j > z$. Thus the net individual income-based welfare effect is $dM + dB +$

[48] As $\kappa(z)$ is negative at low income values, this can be negative.

[49] In the absolute Gini, $\overline{\kappa}(z) = F(z)$.

[50] For the mechanical effect, the tax revenue increase and the individual post-tax income decreases are identical.

$dW = dR \cdot \int_j g_j dj - dz\partial\tau \int_{\{j:z_j \geq z\}} g_j dj$.

The total equality effect is $d\bar{\theta} = d\bar{\theta}_M + d\bar{\theta}_B$:

$$\partial\bar{\theta} = dz\partial\tau \left(-\overline{\kappa}(z)\left[1-F(z)\right] - \kappa(z)\epsilon(z)zf(z)\right)$$

In terms of utility, this affects every individual as $\int_j g_j \frac{\partial U_j}{\partial \bar{\theta}} \cdot \partial\bar{\theta} \cdot dj$. As we assume an homogenous inequality externality and quasi-linearity in consumption such that $\eta = MRS_{x\bar{\theta}} = -\frac{\partial U/\partial\bar{\theta}}{\partial U/\partial x} = -\frac{\partial U}{\partial \bar{\theta}}$, the total welfare effect of the inequality change is $dI = \int_j g_j \cdot (-\eta) \cdot \partial\bar{\theta} \cdot dj = -\eta \cdot \partial\bar{\theta} \cdot \int_j g_j dj$.

The total welfare change, including all channels, is equal to zero at the optimum:

$$dM + dB + dW + dI = 0.$$

Thus, using the expressions for dR and dI, and the expression $\bar{G}(z)\left(1-F(z)\right) = \int_{\{j:z_j \geq z\}} g_j dj / \int_j g_j dj$, we have:

$$\begin{aligned} dz\partial\tau \int_j g_j dj \left[1 - F(z) - f(z)\epsilon(z)z\frac{\tau(z)}{1-\tau(z)}\right] - dz\partial\tau\bar{G}(z)\left(1-F(z)\right)\int_j g_j dj \\ +\eta \cdot \int_j g_j dj \cdot \left[dz\partial\tau\left(\overline{\kappa}(z)\left[1-F(z)\right] + \kappa(z)\epsilon(z)zf(z)\right)\right] = 0 \end{aligned}$$

Dividing by $\int_j g_j dj \cdot dz\partial\tau$ and re-arranging, we find:

$$\frac{\tau(z)}{1-\tau(z)} = \eta \cdot \kappa(z) + \frac{1-F(z)}{z \cdot f(z)} \frac{\left(1 - \bar{G}(z) + \eta\overline{\kappa}(z)\right)}{\epsilon(z)}$$

We use the local Pareto parameter $\alpha(z) = \frac{z \cdot f(z)}{1-F(z)}$ and write $\Upsilon(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\overline{\kappa}(z)$ and find the optimal marginal income tax rates as specified in Equation 6.

C Analytical Solution of the OIT Problem

We write individual utility as;

$$U(x, h, \bar{\theta}) = u(x) - V(h) - \Gamma(\bar{\theta}) \tag{13}$$

where u is the utility of consumption (after-tax income), V is the disutility of work and Γ is the disutility of inequality. Equation (13) assumes that agents are homogeneous, with identical individual utility functions.

At the heart of the model is n, the exogenous wage-earning ability, unobservable to the social planner. There is a continuum of individuals with n varying according to a an exogenous density function $f(n)$, with a cumulative distribution function $F(n)$. Pre-tax earnings are defined as nh, and total consumption is $x = nh - T(nh)$, where $T(\cdot)$ is the tax schedule. The individual maximizes utility by choosing hours worked h given n and $T(\cdot)$. The utility-maximising values of consumption and hours worked are written as

$$x(n), h(n). \tag{14}$$

Given the individual's choice, the social planner chooses the tax schedule to maximize the social welfare function. We assume this to be an additively separable function of individual utility. Accordingly the problem is,

$$\max_{T(\cdot)} \int_{\underline{n}}^{\overline{n}} W(U(x(n), h(n), \bar{\theta}))dF(n). \tag{15}$$

Notice that formulating individual utility as (13) avoids the complication of potentially heterogeneous effects of inequality if the social planner is strictly utilitarian (Benthamite) – in this case only the average inequality externality has an effect. Similarly, a Rawlsian social planner will only take into account the inequality externality on the lowest-utility agent.

The problem (15) is subject to three conditions, the first two of which are standard constraints. First, there is the *revenue constraint* for any required amount R of non-redistributive public goods:

$$R \leq \int_{\underline{n}}^{\overline{n}} T(nh)f(n)dn. \tag{16}$$

For simplicity we assume that $R = 0$.

Second, we have the *incentive-compatibility constraint* from the possibility that an agent with (unobservable) wage-earning ability n could masquerade as an agent with $\hat{n}$. For any person with wage-earning ability n it must be true that:

$$u(x(n)) - V(h(n)) \geq u(x(\hat{n})) - V(h(\hat{n})) \tag{17}$$

where $x(\hat{n})$ and $h(\hat{n})$ are, respectively, the consumption and hours worked if the agent masquerades as someone with ability $\hat{n}$, possibly different from n. The IC constraint (17) ensures that the agent self-selects into the appropriate tax bracket.

Third, we need to introduce the role of inequality into the model. Individuals experience an amount $\bar{\theta}$ of after-tax inequality. This inequality is partly determined by F, the distribution of innate talent, and partly by the choices made by individuals, captured in (14). But it is also partly the result of decisions by the social planner, captured in the tax function T and therefore embedded in (14). We can represent this relationship as the following *inequality condition:*

$$\bar{\theta} = I\left(\boldsymbol{x}, F\right) \tag{18}$$

where $I(\cdot, \cdot)$ is an inequality measure, $\boldsymbol{x}(\cdot)$ is the full set of consumption choices from (14) and $F(\cdot)$ is the distribution function for n.

To complete the model we need an inequality metric $I(\cdot, \cdot)$. We use a specific form of the (absolute) Gini coefficient in after-tax income:

$$I_{\text{Gini}}\left(\boldsymbol{x}, F\right) = \int_{\underline{n}}^{\overline{n}} \kappa(n) x(n) dF(n), \tag{19}$$

where x is after-tax income (consumption), n is the exogenous productivity level, and

$$\kappa(n) = 2F(n) - 1 \tag{20}$$

is an expression for the weight of the agent in the Gini.[51] Expression (19) shows that the absolute Gini can be calculated as a sum of weighted incomes in the population, where the weight $\kappa(n)$ depends only on the *rank* of the agent in the wage-earning ability distribution, which is constant and exogenous by assumption. Using (19), condition (18) becomes

$$\bar{\theta} = \int_{\underline{n}}^{\overline{n}} \left[2F(n) - 1\right] x(n) dF(n).$$

One can also use other inequality metrics based on rank-specific weights, such as those in the Lorenz (Aaberge, 2000) or S-Gini families (Donaldson and Weymark, 1980).

With the inequality externality and inequality metric specified, we note that if the inequality externality $\Gamma(\bar{\theta})$ is linear and we are in a utilitarian framework, the objective function amounts to the SWF derived in Sen (1976) with an additional labor disutility term. This Sen (1976) SWF is also a cumulation of Fehr-Schmidt preferences over the population (Schmidt and Wichardt, 2018), creating another link to the inequality aversion literature.

To solve the analytical problem we first re-write the incentive compatibility constraint. We note that consumption x, i.e. after-tax income, is a function of wage times hours worked: $x = c(nh)$. The individual maximization implies,

$$\frac{dU}{dh} = 0 = u'c'n - V', \tag{21}$$

and from the IC constraint we have (using either the Mirrlees (1971) trick or the envelope condition):

$$\frac{dU}{dn} = u'c'h \tag{22}$$

Taken together these two imply :

[51] This is a slight modification of Equation 27 in Cowell (2000) for the standard (relative) Gini $\int \frac{\kappa'(x(n))x(n)}{\mu(x)} dG(x)$, where $\kappa'(x) = 2G(x) - 1$ is the weight of the agent and $G(x)$ is the CDF of x. If individuals' post-tax income increases with wage-earning ability, the rank-dependent variable $\kappa(n) = \kappa'(x)$. In other words, if there is rank-equivalency between income and ability, we can use the ability ranking to calculate the individual weights in the income inequality metric. Simula and Trannoy (2020), developed simultaneously with this paper, also exploits this rank-invariancy in ability and income. It is a novel method and vastly simplifies the analytical problem.

As we show in Appendix C.I, this assumption is equivalent to assuming that the individuals' second-order conditions hold. For all the numerical simulations we confirm that they in fact do.

$$\frac{dU}{dn} = \frac{V'h}{n} =: g\left(n\right) \tag{23}$$

We can write $T = nh - x$, where x is after-tax consumption.[52] From this and the IC constraint, we observe that the tax schedule implicitly defines both work hours and total individual utility. Instead of setting the tax schedule T, then, we can say that the social planner chooses work hour schedules $h(n)$, utility schedules $U(n)$, and the inequality level $\bar{\theta}$.

The Lagrangian of the full problem classified in Equations 15-13 is,

$$\begin{aligned} L = \int_{\underline{n}}^{\overline{n}} W(U(n))f(n)dn + \lambda(\int_{\underline{n}}^{\overline{n}} \left[nh(n) - x\right] f(n)dn) \\ + \int_{\underline{n}}^{\overline{n}} \alpha(n) \left[\frac{dU}{dn} - g(n)\right] dn + \gamma\left[\bar{\theta} - I_{Gini}\right] \end{aligned} \tag{24}$$

We note that the incentive compatibility constraint can be simplified using integration by parts, and we assume n goes from zero to infinity without loss of generality. After taking these factors into account and combining the rest of the integrals, we have:

$$\begin{aligned} L = \int_{0}^{\infty} \left[W(U(n)) + \lambda(nh(n) - x)\right] f(n) - \alpha(n)g(n) - \alpha'(n)U(n)dn \\ + \alpha(\infty)U(\infty) - \alpha(0)U(0) + \gamma\left[\bar{\theta} - I_{Gini}\right] \end{aligned} \tag{25}$$

We introduce the Gini coefficient in the form,

$$I_{Gini} = \int_{0}^{\infty} \left[2F(n) - 1\right] xf(n)dn = \int_{0}^{\infty} \kappa(n)xf(n)dn \tag{26}$$

Where $f(n)$ and $F(n)$ are the PDF and CDF of n, respectively, and $\kappa(n) = 2F(n) - 1$ is the weight of the agent in the absolute Gini.

The Lagrangian becomes:

$$\begin{aligned} L = \int_{0}^{\infty} \Bigg[\left(W(U(n)) + \lambda\left[nh(n) - x\right] - \gamma\kappa(n)x\right) f(n) - \alpha(n)g(n) \\ - \alpha'(n)U(n)\Bigg] dn + \alpha(\infty)U(\infty) - \alpha(0)U(0) + \gamma\bar{\theta} \end{aligned} \tag{27}$$

From this we can find the first-order conditions with respect to $h(n)$, $U(n)$, and $\bar{\theta}$, as these variables together will implicitly set the tax schedule.[53] Before we begin, note that we can rewrite $x = y(h, U, \bar{\theta}) = u^{-1}(U + V(h) + \Gamma(\bar{\theta}))$, and find expressions for the derivatives y_h, y_U, and $y_{\bar{\theta}}$.[54]

[52] The model is a one-period model and does not contain savings.

[53] We could use the derivative of $x(n)$ instead, but the methods are mathematically equivalent and this procedure is somewhat more straightforward.

[54] Using the rules for derivatives of inverse functions, these expressions are $y_h = \frac{V_h}{u_x}$, $y_{\bar{\theta}} = \frac{\Gamma_{\theta}}{u_x}$, and $y_U = \frac{1}{u_x}$.

The first order conditions are the following:

$$U: \quad 0 = \big[W'(U(n)) - \lambda y_U\big] f(n) - \alpha'(n) - \gamma\kappa(n) f(n) y_U \tag{28}$$

$$h: \quad 0 = \lambda(n - y_h) f(n) - \alpha(n)\frac{V_{hh}h + V_h}{n} - \gamma\kappa(n) f(n) y_h \tag{29}$$

$$\bar{\theta}: \quad 0 = \gamma - \int_0^\infty \gamma\kappa(n) f(n) y_{\bar{\theta}} dn - \int_0^\infty \lambda y_{\bar{\theta}} f(n) dn \tag{30}$$

In the FOC for h we have used that $g = \frac{V_h h}{n}$ from Equation (23), and that $\frac{dg}{dh} = \frac{V_{hh}h + V_h}{n}$. Equation 30 implies,

$$\frac{\gamma}{\lambda} = \frac{\int_0^\infty \frac{\Gamma_{\bar{\theta}}}{u_x} f(n) dn}{1 - \int_0^\infty \frac{\Gamma_{\bar{\theta}}}{u_x}\kappa(n) f(n) dn} \tag{31}$$

Here $\frac{\gamma}{\lambda}$ is the shadow price of the inequality constraint expressed in units of public funds, and $\Gamma_{\bar{\theta}}$ and u_x are derivatives. If $\Gamma(\bar{\theta}) = 0$, as in the standard case when the inequality externality does not exist, then $\gamma = 0$. A negative inequality externality implies a positive $\Gamma_{\bar{\theta}}$, and thus a positive $\frac{\gamma}{\lambda}$. To rephrase, this is the unsurprising result that equality itself has a cost in a world with a negative inequality externality.

Now we move to finding an expression for $\alpha(n)$, the shadow price of the incentive compatibility constraint. We integrate the first order condition for U, Equation 28:[55]

$$\alpha(n) = \int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p))\right] f(p) dp \tag{32}$$

And substitute this into Equation (29):

$$0 = \lambda(n - y_h) f(n) - \gamma\kappa(n) f(n) y_h - \frac{V_{hh}h + V_h}{n}\int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p))\right] f(p) dp \tag{33}$$

$$\frac{(n - y_h)}{y_h} = \frac{\gamma}{\lambda}\kappa(n) + \frac{u_{x(n)}(V_{hh}h + V_h)}{\lambda f(n) n V_h}\int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p))\right] f(p) dp \tag{34}$$

We have that $\frac{n - y_h}{y_h} = \frac{n u_{x(n)}}{V_h} - 1 = \frac{1}{1-t} - 1 = \frac{t}{1-t}$, so we quickly have an expression for optimal marginal tax rates:

$$\frac{t}{1-t} = \frac{\gamma}{\lambda}\kappa(n) + \frac{\zeta_n u_{x(n)}}{\lambda f(n) n}\int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p))\right] dF(p). \tag{35}$$

[55] From the transversality conditions $\frac{dL}{dU(0)} = \alpha(\infty) = 0$. We use the new symbol p to denote the productivity n inside the integral.

Here $\frac{\gamma}{\lambda}$ is the price of inequality in terms of public funds (see Equation 31). If inequality is a negative externality (a public bad), γ will generally be large and positive.[56] The agent's weight in the Gini coefficient, $\kappa(n)$, is negative at the bottom and positive at the top. $\zeta_n = \frac{V_{hh}h}{V_h} + 1$ is a term closely related to the inverse compensated elasticity of labor,[57] and u_x is the marginal utility of consumption.

Two of these terms are equivalent to traditional OIT terms. By denoting the part of the optimal tax function found in Diamond (1998) as $\frac{t_i}{1-t_i}$, we can isolate and evaluate the effect of the inequality externality.

$$\frac{t}{1-t} = \frac{\gamma}{\lambda}\left[\kappa(n) + \frac{\zeta}{f(n)n}\int_n^\infty \frac{u_{x(n)}}{u_{x(p)}}\kappa(p)dF(p)\right] + \frac{t_i}{1-t_i} \tag{36}$$

For clarity let us assume a linear homogeneous inequality externality ($\Gamma(\bar{\theta}) = \eta\bar{\theta}$) and quasi-linearity in consumption.[58] The optimal tax rate condition simplifies to:

$$\frac{t}{1-t} = \eta\kappa(n) + \eta\left(1 + \frac{1}{E_L}\right)\Pi(n)F(n) + \frac{t_i}{1-t_i}, \tag{37}$$

where we denote the distributional thinness measure $\frac{1-F(n)}{f(n)n}$ as $\Pi(n)$.[59] This formula is functionally equivalent to the form we found with the small perturbations method (Equation 6), but uses the exogenous wage n instead of the endogenous earnings z.

C.I. Equivalence of income rankings

In using the modified Gini in Equation 3, we have assumed that the weight of the agent in the ability ranking is the same as the ranking of the agent in the post-tax income ranking. We asserted that this is equivalent to the second-order condition holding, or that $z'(n) > 0$ where $z(n)$ is pretax income (Lollivier and Rochet (1983)). This is not necessarily obvious. Recall that we have a monotonically increasing n: if we have that $x'(n) > 0$, then, we also have the desired equivalence in ability and post-tax rankings. The more standard assumption in the literature is the SOC $z'(n) > 0$. Here we show that $x'(n) > 0$ is equivalent to $z'(n) > 0$.

Assume quasi-linearity for simplicity and define $\mathit{\Omega}(n) = x(n) - V(\frac{z(n)}{n})$. Here $\mathit{\Omega}(n) \geq \mathit{\Omega}(\hat{n})\,\forall\, n, \hat{n}$

[56] If we assume a linear inequality externality of the form $\Gamma(\theta) = \eta\theta$ then $\frac{\gamma}{\lambda} = \eta$ (see Equation 31).

[57] With quasi-linear preferences, $\zeta = \frac{1}{E_L} + 1$.

[58] The resulting utility function is

$$U(x, h, \bar{\theta}) = x - \frac{h^{\left(1+\frac{1}{E_c}\right)}}{\left(1 + \frac{1}{E_c}\right)} - \eta\bar{\theta}$$

Note that with quasi-linearity, $\int_n^\infty \kappa(p)dF(p)$ in (36) simplifies as $\int_n^\infty (2F(n) - 1)\, dF(n) = F(n) - F(n)^2$.

[59] This is the inverse of the local Pareto parameter $\alpha(n)$, which becomes constant in a Pareto distribution. It is also the inverse elasticity of $P(n) = 1 - F(n)$ with regards to n; $\varepsilon_{P,n} = \frac{n}{1-F(n)}\frac{d(1-F)}{dn} = -\frac{nf(n)}{1-F(n)}$.

is equivalent to the IC constraint. The problem becomes

$$max_{V,y} \int \left[\Omega(n) - \Gamma(\bar{\theta})\right] dG(n)$$

$$s.t. \int \left[\Omega(n) + V(\frac{z(n)}{n}) - z(n)\right] dF(n) \leq 0,$$

$$\Omega'(n) = \frac{z(n)}{n^2} V'(\frac{z(n)}{n}),$$

$$\bar{\theta} = I_{Gini},$$

where the second constraint is the individual's FOC. Then we note that:

$$x'(n) = \Omega'(n) + \left(\frac{nz'(n) - z(n)}{n^2}\right) V' = \left(\frac{z(n) + nz'(n) - z(n)}{n^2}\right) V' = \frac{z'(n)}{n} V'$$

And we have the sought-after equivalence; n and $V'(\frac{z(n)}{n})$ are positive, so $z'(n) > 0$ implies $x'(n) > 0$.

Finally, a word of caution: $\frac{t}{1-t}$ can fall below -1 at the bottom of the distribution given a sufficiently large negative externality if everyone works.[60] This is in reality not a solution, as the second-order conditions are violated and the assumption behind the ability-income rank equivalence fails. This example illustrates why our analytical specifications must be taken with caution; in certain settings, and particularly with large externalities, additional constraints should be added. A similar edge case can occur at the top with a large positive externality.

D Additional notes for Section III

D.I. Theoretical ability distributions

We present Rawlsian optimal marginal income tax rates from two theoretical skill distributions in Figure VI, using the Gini as the inequality metric. The first is is a Pareto distribution with $\alpha(n) = 2.0$, which becomes nearly identical to the empirical case at the top of the distribution.[61] The second is a lognormal distribution with $\mu = 2.757$ and $\sigma = 0.5611$, using the values from Mankiw et al. (2009) based on the 2007 U.S. wage distribution.

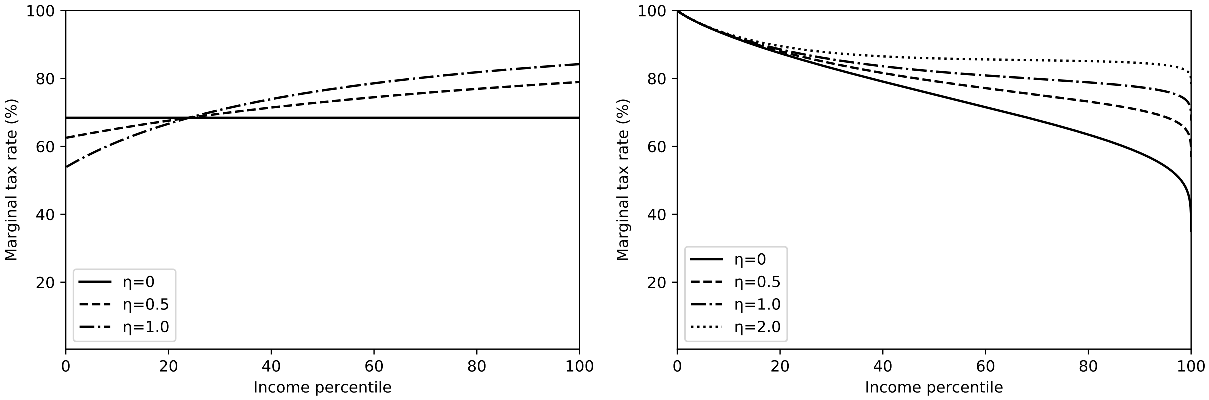
The Pareto case in Figure VIa) illustrates the potentially positive effect of behavioral responses at the bottom. It is socially beneficial for low-income individuals to increase their incomes – so that inequality is reduced – which leads to a small income subsidy at the bottom as compared to the no-externality case. The goal of this tax subsidy is to make individuals internalize that their increased labor supply leads to positive societal outcomes.

[60] The numerical simulations always have an atom of non-working individuals at the bottom to prevent this.

[61] Here $\alpha(n) = 2.0$; in the numerical wage distribution $\alpha(n) = 1.9$. Under this Pareto distribution, second-order conditions fail at the bottom for $\eta = 2.0$. This is therefore not plotted.

Figure VI

Optimal Taxation with Inequality Externalities: Theoretical Ability Distributions



Note: Optimal marginal tax rates for various negative inequality externality magnitudes η. The social planner is Rawlsian and the productivity distribution is (a) a Pareto distribution with $\alpha(n) = 2.0$, (b) a lognormal distribution with $\sigma = 0.39$ and $\mu_{log} = -1$. Inequality aversion estimates indicate $\eta = 1.0$. The solid line, $\eta = 0$, is the standard case of no inequality externality. See Table III for further explanation of the inequality externality magnitudes; $\eta = 2.0$ implies that a representative agent with mean income in a society with Denmark-like income inequality would be indifferent to increasing her income by 25% at the same time as income inequality increased to the United States' current income inequality level. The $\eta = 2.0$ case is excluded from the Pareto simulation because second-order conditions fail at the bottom. The elasticity of labor E_L is 0.3.

The lognormal case further illustrates the localized effects at the top of the distribution. The standard top marginal tax rate in the lognormal case is 0%. With an inequality externality of $\eta = 2.0$ that increases to 67%. This illustrates the Pigouvian correction at the top, and is salient given the local "zero tax at the top"-result of standard models. This local result is not visible in the graph, but is borne out out in the simulations. At the 99th percentile the marginal tax rate increases from 39% in the standard case to 79% when $\eta = 2.0$.

D.II. Varying inequality metrics

In the main specification we used the absolute Gini coefficient for our measure of inequality. Here we explore two different families of inequality metrics. The first is the top income shares also shown in the main text. The second is the S-Gini, which approximates the Gini with a larger focus on either end of the distribution. The distributional weights implied by both families are plotted in Figure VII.[62]

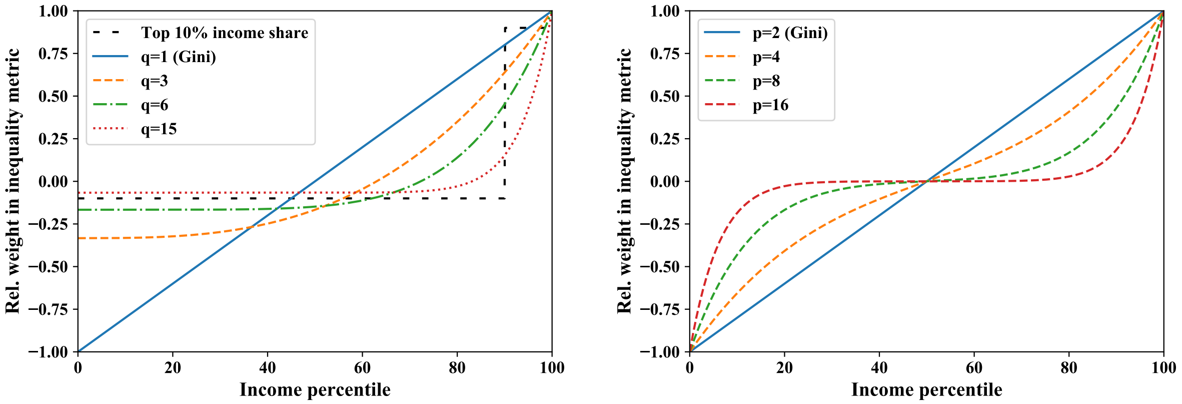
1 Approximating top income shares The first family of inequality metrics, also used in the main robustness test, has some of the properties of top income shares. It is,

$$\bar{\theta} = \int_0^\infty \left[(q+1)F(n)^q - 1\right] x(n) dF(n),\ q \in \mathbb{N}. \tag{38}$$

[62] The weights in Figure VII are normalized such that the top weight is always 1.00. This normalization has no impact on our results due to our re-calculation of η before simulations.

Figure VII

Weights for Families of Inequality Metrics



Note: Consumption weights for inequality metrics used in Appendix D.II. For each individual, their impact on the inequality metric is their proportional weight multiplied by their income. In both figures, the Gini is plotted in solid blue. (a) A family of inequality metrics similar to top income shares, as in Equation 38. The top 10% income share is plotted in dotted black for reference. (b) The S-Gini family from Equation 40.

When $q = 1$, this becomes the absolute Gini coefficient. In all cases, perfect equality implies $\bar{\theta} = 0$ and perfect inequality implies $\bar{\theta} = \mu$ (or $\bar{\theta} = 1$ in the non-absolute family). For increasing q, this indicates an increased focus on the very top of the distribution. The negative externality at the top becomes increasingly concentrated at the very top with increasing q, while the positive externality at the bottom becomes approximately constant for an increasing fraction of the population. In effect, increasing q leads to a metric closer to top income shares, but without the discontinuities that make the analytical problem intractable.

The resulting analytical optimal tax rates with the utility function in 9 become,

$$\frac{t}{1-t} = \eta_q \left[((q+1)F(n)^q - 1) + (1 + \frac{1}{E_c}) \frac{1}{f(n)n} \left[1 - F(n)^q\right] F(n) \right] + \frac{t_{orig}}{1 - t_{orig}}. \tag{39}$$

Here η_q is the magnitude of the inequality externality, which is dependent on q when fitting to empirical data. We ensure that values of η_q are comparable over simulations by re-calculating the parameter from experimental data for each q.[63]

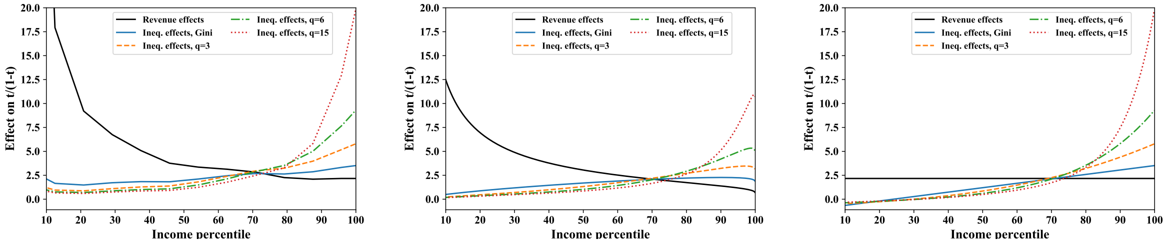
To further illustrate this point, we show the effect of both standard revenue considerations and the new equality considerations on $\frac{t}{1-t}$ with varying inequality metrics in Figure VIII. We present this figure for several different underlying ability distributions. The interaction of equality and revenue considerations can make it difficult to interpret values of t, so this graph illustrates the

[63] We estimated η with data from Carlsson et al. (2005) in the main text. To remain consistent, we have calculated for each inequality metric q comparable η_q from the experimental values in Carlsson et al. (2005) for all following simulations. This means that, while the value of η_q changes, the underlying estimation comes from the same data. This is true for all metrics.

more intuitive impacts on $\frac{t}{1-t}$. All social planners are Rawlsian.[64]

Figure VIII

Effects on $\frac{t}{1-t}$: Top Income Share Externalities



Note: Effects on $\frac{t}{1-t}$ for various negative inequality metrics $\int_0^\infty [(q+1)F(n)^q - 1]\, x(n)dF(n),\ q \in \mathbb{N}$. The social planner is Rawlsian. The magnitude of the inequality externality is in each case calculated as the median value from the empirical inequality aversion estimates in Carlsson et al. (2005). This is done for comparability across inequality metrics. The productivity distribution is (a) the empirical wage distribution from the main text, (b) a log-normal distribution with $\sigma = 0.39$ and $\mu_{log} = -1$, and (c) a Pareto distribution with $a = 2$. See Figure VII for an explanation of the inequality metrics. In particular, larger q indicates that top incomes are increasingly weighted. The elasticity of labor E_L is 0.3.

Several points are worth noting. First, as expected, increasing q leads to a more pronounced effect at the top of the distribution in all cases. Second, below the top the effects of changing the metric are small and generally dampen the effect of the externality. Third, equality considerations are relatively constant over different skill distributions; the major factor changing resulting tax rates over skill distributions are revenue considerations. Fourth, equality considerations are proportionally more important than revenue considerations towards the top of the distribution in all three cases. While by nature dependent on the ability distribution and social welfare function, this last point seems likely to hold in many specifications.

2 The S-Gini The second family of inequality metrics we use is the S-Gini family, which increases the weight of top- and bottom-incomes symmetrically.

$$\bar{\theta} = \int_0^\infty [F(n)^p - (1 - F(n))^p]\, x(n)dF(n),\ p \geq 2. \tag{40}$$

When $p = 2$, this becomes the absolute Gini coefficient. This family also retains the beneficial properties discussed above; perfect equality implies $\bar{\theta} = 0$ and perfect inequality implies $\bar{\theta} = \mu$. For increasing p, the top and bottom is increasingly weighted at the cost of middle incomes. Unlike the previous family, these metrics will always increase if an individual above the median increases their income, as well as decrease if an individual below the median increases their income. The resulting optimal tax rates with the utility function in 9 are,

[64] Equality considerations would not change with any other SWF due to the homogeneous nature of the externality. Revenue effects would decrease at the bottom and converge to the same at the top.

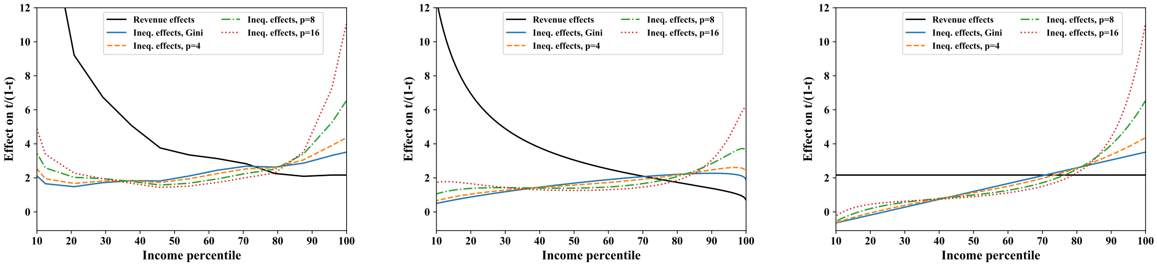
$$\frac{t}{1-t} = \eta_p \left[(F(n)^p - (1-F(n))^p) + (1 + \frac{1}{E_c}) \frac{1}{f(n)n} \nu \right] + \frac{t_{orig}}{1-t_{orig}}, \tag{41}$$

where $\nu = \frac{1}{p+1}\left[1 - [F(n)^{p+1} + (1-F(n))^{p+1}]\right]$.

In Figure IX we show the effect of changing p on $\frac{t}{1-t}$ with the same methodology as in Figure VIII. Increasing p again leads to larger effects towards the top of the distribution and relatively small changes at the bottom. It is notable that the effects at the bottom remain small despite the increased magnitude of the positive externality on these individuals' income. This is driven by the opposition of the mechanical and behavioral channels discussed in the main text. Both equality effects – the internalization of the externality and the increased want for equality – move in the same direction at the top, but work against each other near the bottom.

Figure IX

Effects on $\frac{t}{1-t}$: The S-Gini Family



Note: Effects on $\frac{t}{1-t}$ for various S-Ginis. The social planner is Rawlsian. The magnitude of the inequality externality is held constant for all p at the upper bound of the median value from the empirical inequality aversion estimates in Carlsson et al. (2005). The productivity distribution is (a) the empirical wage distribution from the main text, (b) a log-normal distribution with $\sigma = 0.39$ and $\mu_{log} = -1$, and (c) a Pareto distribution with $a = 2$. See Figure VII for an explanation of the inequality metrics. In particular, larger p indicates that top and bottom income variation is weighted more than middle-income variation. The elasticity of labor E_L is 0.3.

The majority of the new insight noted in the previous subsection also hold for the S-Gini. Unlike in the top income shares, however, the benefits of taxing near the bottom increase with increasing p. This is a somewhat surprising result. It is due to the mechanical effect being more potent when bottom externalities are very large; in effect, the average inequality metric weight above increases rapidly near the bottom. This leads to the generally large equality benefits from the mechanical effect being even larger than the increased benefits of subsidizing the poor to work more. We caution that this is a particularly model-driven result.

A last caveat; throughout the paper we use a family of *absolute* inequality metrics. This is done to keep scale independence in the additive utility function. However, as this means that the inequality metric can increase without bounds, caution is required when working with large externality values. A further exploration of other functional forms would be beneficial to understand how this changes the optimal tax problem.

D.III. A squared inequality externality function

Our framework is sufficiently general for other functional forms of the MRS, or equivalently $\Gamma(\bar{\theta})$, the inequality function from the utility function (see Appendix C). Let us use $\Gamma(\bar{\theta}) = \eta(\bar{\theta} - \bar{\theta}_{opt})^2$, such that:

$$U(x, h, \bar{\theta}) = x - \frac{h^{\left(1+\frac{1}{E_c}\right)}}{\left(1 + \frac{1}{E_c}\right)} - \eta(\bar{\theta} - \bar{\theta}_{opt})^2 \tag{42}$$

The resulting analytical optimal tax rates are:

$$\frac{t}{1-t} = 2\eta(\bar{\theta} - \bar{\theta}_{opt}) \left[\kappa(n) + \frac{\zeta}{f(n)n} \int_n^{\infty} \kappa(p) f(p) dp\right] + \frac{t_{orig}}{1 - t_{orig}} \tag{43}$$

Comparing these tax rates to Equation 36, we see that the effect of the inequality externality is attenuated by a factor of $2(\bar{\theta} - \bar{\theta}_{opt})$. The policy effect of the inequality externality will be larger in societies with high after-tax inequality. We find this intuitive; tax systems responding to inequality will respond more when initial inequality is high. The result is the same when using the small perturbations method.

Also note that this solution is endogenous, as $\bar{\theta}$ depends on the tax schedule. We thus need numerical methods to solve for the optimal tax schedule. This is not a unique feature of this formulation, and also occurs when the social weights are endogenous as in the non-Rawlsian solutions.

We do not perform numerical simulations in this case, primarily because of the complicated nature of estimating a suitable η when we have another unknown variable in $\bar{\theta}_{opt}$.

References

Aaberge, R. (2000). Characterizations of lorenz curves and income distributions. *Social Choice and Welfare*, 17(4):639–653.

Aaberge, R. and Colombino, U. (2013). Using a microeconometric model of household labour supply to design optimal income taxes. *The Scandinavian Journal of Economics*, 115(2):449–475.

Alesina, A. and Giuliano, P. (2011). Preferences for redistribution. In *Handbook of social economics*, volume 1, pages 93–131. Elsevier.

Anbarci, N., Escaleras, M., and Register, C. A. (2009). Traffic fatalities: does income inequality create an externality? *Canadian Journal of Economics/Revue canadienne d'économique*, 42(1):244–266.

Aronsson, T. and Johansson-Stenman, O. (2008). When the Joneses' consumption hurts: Optimal public good provision and nonlinear income taxation. *Journal of Public Economics*, 92(5-6):986–997.

Aronsson, T. and Johansson-Stenman, O. (2010). Positional concerns in an OLG model: optimal labor and capital income taxation. *International Economic Review*, 51(4):1071–1095.

Aronsson, T. and Johansson-Stenman, O. (2015). Keeping up with the Joneses, the Smiths and the Tanakas: on international tax coordination and social comparisons. *Journal of Public Economics*, 131:71–86.

Aronsson, T. and Johansson-Stenman, O. (2016). Inequality aversion and marginal income taxation. *Working Papers in Economics*, 673.

Aronsson, T. and Johansson-Stenman, O. (2020). Optimal second-best taxation when individuals have social preferences. *Working Paper*.

Ashworth, J., Heyndels, B., and Smolders, C. (2002). Redistribution as a local public good: an empirical test for Flemish municipalities. *Kyklos*, 55(1):27–56.

Atkinson, A. B. and Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, 6(1-2):55–75.

Benabou, R. and Ok, E. A. (2001). Social mobility and the demand for redistribution: the POUM hypothesis. *The Quarterly Journal of Economics*, 116(2):447–487.

Bergh, A., Nilsson, T., and Waldenström, D. (2016). *Sick of Inequality?: An Introduction to the Relationship Between Inequality and Health*. Edward Elgar Publishing.

Blundell, R. and Shephard, A. (2011). Employment, hours of work and the optimal taxation of low-income families. *The Review of Economic Studies*, 79(2):481–510.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Bonica, A., McCarty, N., Poole, K. T., and Rosenthal, H. (2013). Why hasn't democracy slowed rising inequality? *Journal of Economic Perspectives*, 27(3):103–24.

Boskin, M. J. and Sheshinski, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *The Quarterly Journal of Economics*, pages 589–601.

Burgoon, B., van Noort, S., Rooduijn, M., and Underhill, G. R. (2018). Radical right populism and the role of positional deprivation and inequality. Technical report, LIS Working Paper Series.

Carlsson, F., Daruvala, D., and Johansson-Stenman, O. (2005). Are people inequality-averse, or just risk-averse? *Economica*, 72(287):375–396.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869.

Cingano, F. (2014). Trends in income inequality and its impact on economic growth.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic literature*, 46(1):95–144.

Cooper, D. J. and Kagel, J. (2016). Other-regarding preferences. *The handbook of experimental economics*, 2:217.

Cowell, F. A. (2000). Measurement of inequality. *Handbook of income distribution*, 1:87–166.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348–361.

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

Diamond, P. A. and Mirrlees, J. A. (1971). Optimal taxation and public production II: Tax rules. *The American Economic Review*, 61(3):261–278.

Donaldson, D. and Weymark, J. A. (1980). A single-parameter generalization of the gini indices of inequality. *Journal of economic Theory*, 22(1):67–86.

Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., and Sobel, J. (2011). Other-regarding preferences in general equilibrium. *The Review of Economic Studies*, 78(2):613–639.

Fairbrother, M. and Martin, I. W. (2013). Does inequality erode social trust? Results from multilevel models of US states and counties. *Social science research*, 42(2):347–360.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Flood, S., King, M., Rodgers, R., Ruggles, S., and Warren, J. R. (2018). Integrated public use microdata series, current population survey: Version 6.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V6.0.

Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status.* Oxford University Press.

Kanbur, R., Keen, M., and Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income-poverty. *European Economic Review*, 38(8):1613–1632.

Kanbur, R. and Tuomala, M. (2013). Relativity, inequality, and optimal nonlinear income taxation. *International Economic Review*, 54(4):1199–1217.

Laffer, A. B. (2004). The laffer curve: Past, present, and future. *Backgrounder*, 1765:1–16.

Lindbeck, A. (1985). Redistribution policy and the expansion of the public sector. *Journal of Public Economics*, 28(3):309–328.

Lockwood, B. B., Nathanson, C. G., and Weyl, E. G. (2017). Taxation and the allocation of talent. *Journal of Political Economy*, 125(5):1635–1682.

Lollivier, S. and Rochet, J.-C. (1983). Bunching and second-order conditions: A note on optimal tax theory. *Journal of Economic Theory*, 31(2):392–400.

Mankiw, N. G., Weinzierl, M., and Yagan, D. (2009). Optimal taxation in theory and practice. *Journal of Economic Perspectives*, 23(4):147–74.

Manning, A. et al. (2015). Top rate of income tax. *Centre for Economic.*

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of public Economics*, 6(4):327–358.

Oswald, A. J. (1983). Altruism, jealousy and the theory of optimal non-linear taxation. *Journal of Public Economics*, 20(1):77–87.

Pastor, L. and Veronesi, P. (2018). Inequality aversion, populism, and the backlash against globalization. Technical report, National Bureau of Economic Research.

Pauly, M. V. et al. (1973). Income redistribution as a local public good. *Journal of Public economics*, 2(1):35–58.

Persson, M. (1995). Why are taxes so high in egalitarian societies? *The Scandinavian Journal of Economics*, pages 569–580.

Piketty, T. and Saez, E. (2007). How progressive is the US federal tax system? a historical and international perspective. *Journal of Economic perspectives*, 21(1):3–24.

Piketty, T. and Saez, E. (2013). Optimal labor income taxation. In *Handbook of Public Economics*, volume 5, pages 391–474. Elsevier.

Piketty, T., Saez, E., and Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–71.

Prete, V., Sommacal, A., Zoli, C., et al. (2016). Optimal non-welfarist income taxation for inequality and polarization reduction. Technical report.

Rothschild, C. and Scheuer, F. (2016). Optimal taxation with rent-seeking. *The Review of Economic Studies*, 83(3):1225–1262.

Rueda, D. and Stegmueller, D. (2016). The externalities of inequality: Fear of crime and preferences for redistribution in western europe. *American Journal of Political Science*, 60(2):472–489.

Rufrancos, H., Power, M., Pickett, K. E., and Wilkinson, R. (2013). Income inequality and crime: A review and explanation of the timeâ series evidence. *Sociology and Criminology-Open Access.*

Sadka, E. (1976). On income distribution, incentive effects and optimal income taxation. *The review of economic studies*, 43(2):261–267.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1):205–229.

Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, 106(1):24–45.

Sandmo, A. (1975). Optimal taxation in the presence of externalities. *The Swedish Journal of Economics*, pages 86–98.

Schmidt, U. and Wichardt, P. C. (2018). Inequity aversion, welfare measurement and the gini index.

Seade, J. K. (1977). On the shape of optimal tax schedules. *Journal of public Economics*, 7(2):203–235.

Sen, A. (1976). Real national income. *The Review of Economic Studies*, 43(1):19–39.

Simula, L. and Trannoy, A. (2020). Gini and optimal income taxation by rank. *CESifo Working Paper.*

Stiglitz, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of public economics*, 17(2):213–240.

Thurow, L. C. (1971). The income distribution as a pure public good. *The Quarterly Journal of Economics*, pages 327–336.

Tuomala, M. et al. (1990). Optimal income tax and redistribution. *OUP Catalogue*.

You, J.-s. (2014). Land reform, inequality, and corruption: A comparative historical study of Korea, Taiwan, and the Philippines.