INEQUALITY AS AN EXTERNALITY: CONSEQUENCES FOR TAX DESIGN*

MORTEN NYBORG STØSTAD† AND FRANK COWELL‡

Abstract

This paper proposes to treat inequality as an economic externality in order to introduce the societal effects of inequality into welfarist models. These effects can include (but are not limited to) changes in political efficiency, economic growth rates, or interpersonal trust levels. 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 individualist parameters. Novel policy implications are illustrated through the classical optimal non-linear income taxation model, where an income inequality externality leads to a trade-off between standard revenue effects and equality effects. Policy consequences are disproportionately located at the top, where optimal marginal tax rates are strongly and robustly dependent on the magnitude and direction of the inequality externality. We use several real-world examples to show that tax policy previously unsupported by optimal taxation theory can be explained by an appropriate inequality externality. The findings indicate that the magnitude of the inequality externality should be considered a crucial economic variable. *JEL* Codes: H21, H23, D62, D63

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

†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 inequality itself affects society and thus individuals.[1] 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, and 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 is *itself* policy relevant.

Despite this, and despite a large empirical academic literature attempting to explore such effects,[2] inequality itself does not affect individuals in the vast majority of economic models. While there are consequences of inequality in standard welfarist frameworks – usually the cumulative effect of diminishing marginal utilities of income and inequality-averse social welfare functions – agents are assumed to be individually indifferent to changes in societal inequality of any kind. This holds true in nearly all welfarist models.

This paper removes the need for this assumption through a simple and generalizable framework. Further, it shows that such an inequality externality can have significant consequences for optimal public policy. Simply put, including an inequality externality in the analysis tends to drastically change policy advice.

To illustrate we use the example of an income inequality externality in the optimal income taxation. 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. The novel effects are large, intuitively appealing, and reflect how both policy leaders and economists describe policy goals. We find several examples of real-world policies that cannot be rationally justified under standard optimal taxation rules that are welfare-optimizing when inequality is an externality.

The insights we present for optimal taxation can be extrapolated to a wide array of economic models. In general, the key assumption that allows our approach in the standard optimal taxation (OIT) literature is the assumption of individualist utility functions. Implicitly, this assumption is equivalent to stating that any societal effects of inequality are negligible for every individual at every point of the income distribution.[3] This implies that every individual is indifferent between living in a perfectly equal and perfectly unequal society.

This assumption seems overly ambitious, particularly when the analyzed policy directly affects

[1] See, for example, Plato (est. 360 B.C.); Aristotle (est. 350 B.C.); de Tocqueville (1835); Keynes (1919); Roosevelt (1939); Sen (2001); OECD (2006); Obama (2011); Johnson (2013); Pope Francis (2014); Sanders (2014); Lagarde (2014); Greenspan (2014); Draghi (2017); and Wolf (2019), among many other examples. Full quotes are presented in Appendix A.

[2] 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).

[3] Or, alternatively, that the cumulative positive inequality effects are exactly utility-equal to the cumulative negative inequality effect for all agents.

the income distribution. Our results indicate that it does indeed have large optimal policy consequences. While this paper concerns itself solely with the OIT model, we also question which other standard economic models would be similarly impacted, as our framework is general enough to be extended to many theoretical models. To illustrate our method, we 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 societal inequality even absent any other-regarding preferences. Even a perfectly self-centered individual may care a great deal about the *effects* of inequality. This implies that 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 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.

We formalize our model by analyzing income inequality itself as an economic externality. Our rationale for this formulation is the following. First, individual labor supply decisions affect income inequality, but workers are not incentivized to take their effect on inequality into account when making labor supply decisions. Second, income inequality affects the utility of other individuals, potentially through indirect channels, 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 remained silent on Thurow's idea. The two 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.[5] We go further and create a more extensive framework for discussing these ideas,

[4]This argument is similar to part of the argument for why consumption enters the utility function. We often value what consumption will give us – such as the act of eating, good health, or being entertained – and not necessarily the act of consumption or buying goods itself. We also note that our formal model does not require agents to be aware of or estimate the inequality externality.

[5]Other notable and related concepts are the status concerns discussed in Frank (1985) and the relative income

including micro-founding a large set of potential inequality effects that do not necessarily only affect consumption. Overall, this work aims to revitalize this small theoretical literature and create a simple and generalizable framework around inequality's *three welfarist consequences*; unequal marginal utilities of income, welfare-loss from differing social weights, and 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; the diminishing marginal utility of income and social welfare weights (potentially generalizable, Saez and Stantcheva (2016)). These represent the differing value of a dollar for different agents and philosophical fairness concerns by the social planner, respectively. 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.[6] The non-linear 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. This leads to novel policy conclusions. While agent behavior remains the same, the social planner's incentives for taxation changes. This changes the model outcomes. 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). With an inequality externality, these effects do not only change the revenue collected but also equality itself. This has a pertinent welfare effect. The comparison to Saez (2001) emphasizes that our model utilizes conventional methods while introducing a key welfare impact of taxation that has been traditionally neglected.

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

concerns explored in Clark et al. (2008).

[6] 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.

the social planner's incentive to change incomes will be larger towards the ends of the distribution.[7] 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 bottom marginal tax rates, where marginal tax rate changes affect the whole distribution.

Our numerical simulations support the 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 at the top of the distribution. We observe values between 37% and 92% under reasonable externality values, holding all other parameters constant. 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, even if the social planner is Rawlsian).

As the Mirrlees model has been extensively used to discuss optimal top tax rates – see Saez (2001) and Piketty et al. (2014), for example – this finding has special significance. Furthermore, it indicates that an understanding of inequality's externality effects could be particularly important to understand policy disagreements on top tax rates.

Other theoretical results in our model 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 find optimal top tax rates above the revenue-maximizing Laffer rate, as direct equality effects imply that tax rates above the revenue-maximizing rate can be optimal. 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 welfare consequences of inequality* and build a framework to describe the causes and formulations

[7]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.

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 (Saez and Stantcheva, 2016). 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 B.

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.[8]

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 B.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.[9]

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

Table I

The Three Welfarist Consequences of Inequality

	Diminishing marginal utility of income	Generalized social weights	Inequality externality
Formulation	$\int_i g_i \boldsymbol{U_i}(\boldsymbol{x_i}, \bar{\theta}, ...)di$	$\int_i \boldsymbol{g_i} U_i(x_i, \bar{\theta}, ...)di$	$\int_i g_i U_i(x_i, \boldsymbol{\bar{\theta}}, ...)di$
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.

[8] 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.

[9] 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 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 $\bar{P} = I(\boldsymbol{O})$, 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, increasing $\bar{P}$ and affecting $U_i(...)$. Thus, inequality leads to more pronounced political polarization and subsequent individual utility impacts.[10]

- 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,

[10] The same argument also holds for diversity of opinions more generally. A generous view is that increased income inequality would 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.[11] 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 $\overline{G} = \sum_i \sum_j (x_i - x_j)$. If $\overline{G}$ 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

[11] 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, education levels, individual freedoms, the distribution of power, 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 $x_i(\bar{\theta})$.[12] 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.[13] It is also unnecessary: the separate contributions of ORP and inequality effects 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.

- Equality becomes policy-relevant in itself.

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

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

- Individuals make a socially suboptimal work decision. With a negative income inequality 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 ($\frac{\partial \theta}{\partial x_i} > 0$)[14] 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.[15] 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 B for further discussion.

[14]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 $\frac{\partial I_{\text{Gini}}}{\partial x_i} = \kappa(x_i)$.

[15]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.[16] The primary cost-benefit trade-off in this model has been between revenue redistribution and the efficiency losses from taxation. Traditionally, income inequality has not had an effect on the individual. Following the model, optimal tax-policy disagreements have largely focused on three principal areas: (i) the extent of behavioral responses to taxation, (ii) the choice of a social welfare function, and (iii) the shape of the current wage-earning ability distribution.

In addition, works since at least Sandmo (1975) have examined the effect of various externalities on the optimal tax problem. Aronsson and Johansson-Stenman (2016) and Aronsson and Johansson-Stenman (2020) are closest to our analysis, though neither expressly considers inequality as an externality. 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, focus on ORP rather than on the overall effects of inequality, and discuss a set of fully-specified utility functions. 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. Relative income concerns have been modelled by 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.

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 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, regardless of origin. This naturally increases the role of income taxation in the optimal policy solution.

The goal of this work is to suggest another core dimension to the optimal income tax problem.

[16] 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).

This dimension is the extent to which income inequality is an externality. It is the first work to use a continuum of agents in the inequality-centered OIT problem, and solves the problem with both the small perturbations method of Saez (2001) and under a full analytical specification. In addition, it is one of a very small subset of papers considering a non-linear and comprehensive post-tax income externality of any type in the continuous problem.

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 corresponds to introducing an 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.

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' after-tax income.[17] 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. However, their actions have welfare-pertinent effects as the change in income inequality impacts n other agents.[18]

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 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. In a quasi-linear utility function, which we use in the small perturbation solution in Appendix C, these assumptions imply a linear additive income inequality externality.[19] The more general form is found in the analytical solution of the problem in Appendix D.

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

$$\bar{\theta}_{\text{Gini}}\left(\boldsymbol{x}, F\right) = \int_0^\infty \kappa(z)x(z)dF(z), \tag{3}$$

[17] 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.

[18] 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.

[19] Mathematically, $U(x, h, \bar{\theta}) = x - v(h) - \eta\bar{\theta}$. If the inequality 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 E.III).

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

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

is the weight of the agent in the Gini (Cowell, 2000). Expression (3) shows that the absolute Gini can be calculated as a sum of weighted incomes in the population, where the weight $\kappa(z)$ depends only on the *rank* of the agent in the income (or earnings) distribution.

This specification streamlines the analytical problem. One can also use other inequality metrics based on 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). We extend our analysis to the S-Gini and a generalised Gini family approximating top income shares in Appendix E.II. Absolute inequality metrics are used to keep scale invariance.

2 Optimal marginal tax schedule To calculate the optimal income taxation results we use the small perturbations method from Saez (2001). The solution to the full analytical approach – which is more general – is presented in Appendix D. We assume no income effects and that inequality affects everyone equally (a *homogeneous* inequality externality).

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

$$\tau(z) = \frac{1 + \Upsilon(z) - \bar{G}(z)}{1 + \Upsilon(z) + \alpha(z)\epsilon(z) - \bar{G}(z)}. \tag{5}$$

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 public bad. If it is negative, inequality is 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 use several of the standard parameters from the optimal taxation literature: the local Pareto parameter $\alpha(z)$, the elasticity of earnings $\epsilon(z)$ (with respect to $1 - \tau(z)$), and the average social welfare weight above $\bar{G}(z)$.[20]

At the very top of the income distribution, this simplifies to:

$$T'(z) = \frac{\eta\kappa(z)}{1 + \eta\kappa(z)}. \tag{6}$$

The intuition of these two expressions is as follows. There are two localized effects of a marginal tax raise on post-tax income, and thus two effects of a marginal tax raise on post-tax income inequality.[21] These two terms correspond to the behavioral responses and the mechanical effect in

[20] $\alpha(z) = \frac{zf(z)}{1-F(z)}$ is a distributional measure which becomes constant in a Pareto distribution. See Saez (2001) for further discussion.

[21] 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. Such flat income changes affects the trade-off between individual income and income inequality by making individual income of relatively larger

the classical OIT literature, both of which have relevant equality effects. To further discuss the intuition behind these terms, we now discuss each separately.

First term: 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 the tax bracket $T'(z)$. These agents work less due to the tax increase. The behavioral responses decreases the income of these individuals, which affects inequality. This behavioral benefit remains even if the individual is at the very top of the income distribution, and corresponds to a Pigouvian tax designed to correct the individual's socially suboptimal labor decision. This implies a potentially positive equality effect of the behavioral responses, which are unambiguously negative in the standard framework.[22]

This term is best understood as an attempt at internalizing individuals' socially suboptimal labor choice. The suboptimality of this choice differs in magnitude and direction based on the position of the individual. As an example, if the agent is at the top ($\kappa(z) < 0$) in a negative inequality externality framework ($\eta > 0$), their unbiased labor choice is too large, skewed towards increasing individual income at a social cost. This term internalizes this social cost or benefit.

The term is affected by four parameters. How the agent affects inequality, represented by their weight in the inequality metric $\kappa(z)$. How inequality affects other agents, represented by the externality magnitude η. The degree to which agents substitute away from work when taxed, represented by the elasticity $\epsilon(z)$. And finally the total amount of agents at the tax bracket z, represented by the distributional term $\alpha(z)$.

If individuals contribute more to the inequality externality, i.e. the absolute value of their respective $\kappa(n)$ is high, the subsidy or tax will be proportionally larger. Subsequently, this term is large at the ends of the distribution.

This Pigouvian term invalidates three classic results from the literature based on Mirrlees (1971).[23] Specifically:

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., 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.
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. Thus we can also have a tax subsidy at the bottom of the income distribution.
3. Seade (1977) argued that the optimal marginal tax rate should be between zero and one. Given the above result for the bottom of the income distribution this is no longer true – one can have

importance without changing the inequality metric.

[22] 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.

[23] The original results are fragile, and change with many small modifications to the model. See Stiglitz (1982) and Saez (2001) for examples.

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. The result has been controversial; here we argue that it shows an intrinsic limitation of the Mirrlees (1971) model. In the classic model inequality in itself is not valued by individuals, who are individually indifferent if inequality changes. As such, the income of the top agent is irrelevant for society unless it can directly contribute to 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.

Second term: An increased taste for equality The second term is from the mechanical effect, which denotes the average tax increase on every individual above the tax bracket in question. Due to the assumption of no income effects, these agents do not change their labor choice. However, their average tax rate increases, and thus the income of each agent above the tax bracket decreases (each by the same amount). If the revenue is redistributed equally, absolute inequality decreases by definition as long as the tax perturbation is not at the very top, where there are no agents above, or at the very bottom, where every agent is above. In every other case, this leads to a pertinent welfare change.

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 $\alpha(z)$).

This term indicates the increased social willingness to change inequality levels by raising or lowering average tax rates on all agents above a given tax bracket. It has the same sign as the inequality externality η, as $\bar{\kappa}(z) = F(z)$ for the absolute Gini coefficient. Indeed, $\bar{\kappa}(z)$ is always positive for all inequality metrics with monotonically increasing weights. 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 internalizes the externality, which impacts $\tau(z)$ proportionately to the individual's weight in the inequality metric (negative at the bottom, positive at the top) and the direction and magnitude of the externality. The second term symbolizes the increased social willingness to pay for equality, and changes $\tau(z)$ proportionately to the direction and magnitude of the externality.

Both new terms always change after-tax income inequality in the direction of the externality. Thus, 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 but significant 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. Additionally, optimal marginal tax rates above the median wage always increase (decrease) as compared to the standard case given a negative (positive) inequality externality.

III.A Numerical Simulations

We use numerical calculations to find optimal tax rates in the presence of an inequality externality. This section uses wage distribution parameters instead of income distribution parameters, as the former is assumed exogenous and the latter is tax-endogenous. This simplifies the empirical application while retaining the structure and intuition developed in the previous section. See the analytical solution in Appendix D for the wage-specific calculation. The main focus of the numerical simulations will be on the effect of the inequality externality.[24]

Method In the traditional optimal tax literature, tax rates are largely determined by three factors; labor (or earnings) elasticities, the social welfare function, and the shape of the wage-earning ability distribution (see Mankiw et al. (2009)). The first, the labor elasticities, will be kept constant for simplicity. We assume that the elasticity of labor supply is constant at $E_L = 0.3$ for all income levels, an empirically reasonable mid-range value.[25]

We use two types of social welfare functions; (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 Utilitarian-like social welfare function, $W = \beta e^{-U/\beta}$. The parameter β indicates the scaling of the social weights. When $\beta \to 0$ we approach the Rawlsian case. We set $\beta = 10$, indicating that the social planner dislikes inequality somewhat more than the strict Utilitarian.

For the shape of the wage-earning ability distribution $F(n)$, our main specification in Figure I uses empirical survey data for the 2018 U.S. wage distribution gathered from the Annual Social and Economic Supplement of the Current Population Survey.[26] Because survey data is incomplete towards the top, we assume that the wage distribution approximates a Pareto distribution for the top 5% with the Pareto parameter $\alpha(n) = 2.0$, as in Saez (2001). This is consistent with the values we find in the empirical wage distribution close to the top 5%. In addition to Figure I, we also present two standard theoretical wage distributions in Appendix E.

We use the following optimal marginal tax rates throughout;

$$\tau(n) = \frac{1 + \Upsilon(n) - \bar{G}(n)}{1 + \Upsilon(n) + \alpha(n)\epsilon_L(n) - \bar{G}(n)}. \tag{7}$$

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

[25] Note that the earnings elasticity with respect to $1 - t$, denoted $\epsilon(z)$, and the elasticity of labor supply, denoted E_L, are related in this model as $\epsilon(z) = E_L \big/ (1 + E_L)$.

[26] 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.

Where $\Upsilon(z) = \eta\alpha(n)\epsilon(n)\kappa(n) + \eta\bar{\kappa}(n)$. This is equal to our earlier result, but under pre-tax wage n instead of pre-tax income z (see Appendix D).[27]

We use a range of realistic values of η to illustrate the potential tax policy consequences given various magnitude of the income inequality externality. To find such a range of η we present an estimate 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 trade-off their hypothetical grandchildren would prefer, allows us to find individual inequality aversion determined to an interval. We use these data to estimate a distribution of η.[28]

The median value in the survey is approximately $\eta = 1.00$, and a majority of respondents have $0.26 < \eta < 2.18$. A negative η – indicating a preference for inequality, or that inequality is a positive externality – is only observed in 7% of respondents.[29]

As these numbers are rather abstract, we present an alternative way of understanding η through equivalent incomes. Answering the following question pins down η:[30] *What multiple of their current income should an average agent require to move from Denmark-like to United States-like inequality?*

Answering the question creates equivalent incomes for differing inequality levels. These equivalent incomes for Denmark and the United States, and their corresponding η, are shown in Table II. As an example, if we have an inequality externality of $\eta = 1.0$, 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 $\eta = 0$, the agent is indifferent without any change to their income. The change in income compensates for 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.

Based on these two techniques we use the range $-0.5 \leq \eta \leq 2.0$ for the inequality 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),

[27] This implies the following utility function, identical under logarithmic transformations:

$$U(x, h, \bar{\theta}) = x - \frac{h^{\left(1+\frac{1}{E_c}\right)}}{\left(1 + \frac{1}{E_c}\right)} - \eta\bar{\theta} \tag{8}$$

[28] 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.

[29] The frequency distribution is the following (cumulative frequency in parenthesis): $-\infty < \eta \leq -0.89 = 0.03\,(0.03)$, $-0.89 < \eta \leq 0.00 = 0.04\,(0.07)$, $0.00 < \eta \leq 0.26 = 0.11\,(0.18)$, $0.26 < \eta \leq 0.51 = 0.11\,(0.29)$, $0.51 < \eta \leq 1.11 = 0.23\,(0.52)$, $1.11 < \eta \leq 1.56 = 0.21\,(0.73)$, $1.56 < \eta \leq 2.18 = 0.11\,(0.84)$, $2.18 < \eta \leq 2.53 = 0.07\,(0.90)$, $2.53 < \eta \leq 3.00 = 0.04\,(0.94)$, $3.00 < \eta = 0.06\,(1.00)$. Due to the design of the experiment, any one individual's inequality aversion is only pinned down to a range.

[30] 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.

Table II

The Magnitude of Inequality Externalities η

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

Note: Which multiple of their current income μ 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 from the utility function in Equation 8.

and second we numerically ensure that the incentive compatibility constraint is satisfied for every agent. We also perform numerical small-perturbation checks to ensure local optimality of the solution in the theoretical skill distributions in Appendix E.I. The success of these checks indicate that the optimal tax rate solution is generally robust.[31]

Results Our main specifications are presented in Figures I. The introduction of even a small inequality externality substantially changes the optimal tax structure. The effect is larger towards the top of the income distribution.

The top marginal tax rate increases from 68% to 89% when assuming a moderately large negative inequality externality, $\eta = 2.0$ (see Table II). With a small positive inequality externality ($\eta = -0.5$), the optimal top marginal tax rate is only 37%.

The effects of the externality are concentrated at the top in all specifications. This remains true with other social welfare functions (not shown) and is even more pronounced for other inequality metrics (shown in Appendix E.II). This finding is due to how easily the social planner can target top income externalities by changing marginal taxes near the top. To explain, note that the externalities with largest magnitudes (in proportion to agent incomes) are at the top and bottom of the distribution. However, bottom marginal tax changes affect everyone; thus the social planner cannot specifically target only low-income externalities. Top marginal tax rates target only top incomes, and are thus relatively more effective.

At the bottom of the distribution, all Rawlsian cases converge to a very high marginal tax rate. This is due to the large positive mechanical revenue effects of increasing bottom marginal tax rates.[32] In the semi-Utilitarian case, these effects are still large but do not fully dominate.

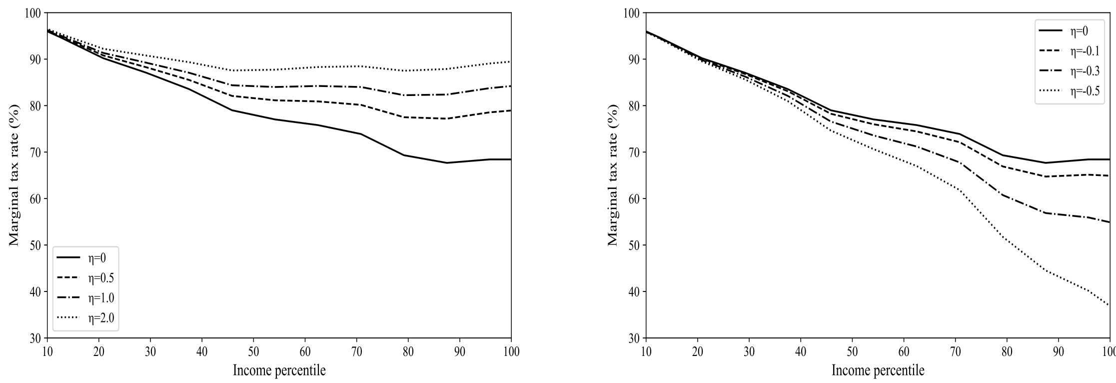
In terms of the effect of the inequality externality, the main difference between the Rawlsian and semi-Utilitarian simulations are the larger externality effects at the lower ends of the distribution

[31] The small perturbation check is only possible on the theoretical skill distributions due to the low granularity of the empirical data. The small perturbations are individual increases and decreases of one per cent from the optimum at each wage-earning level in the skill distribution (or, equivalently, at each income level). For each individual perturbation we re-compute all labor supply choices, all consumption bundles, and all tax revenues, before a check is performed to ensure that the total social welfare decreases from the optimal solution if second-order conditions still hold.

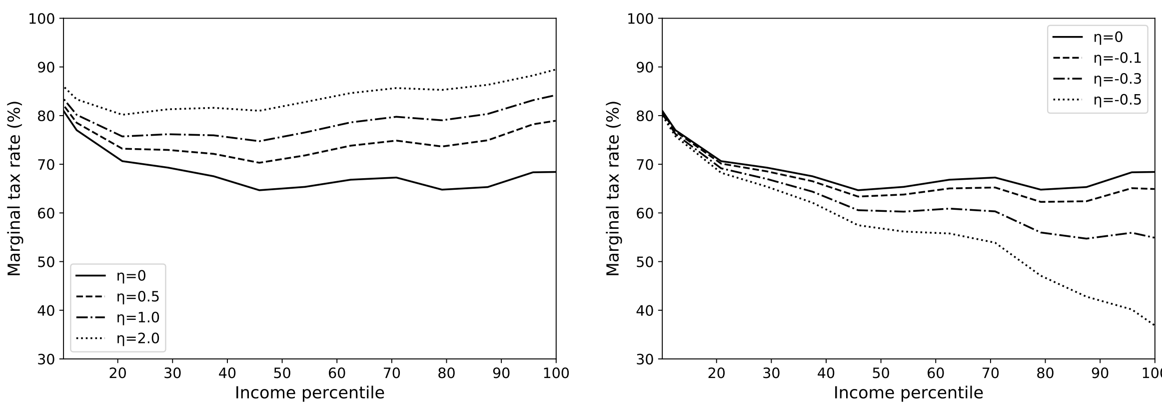
[32] Increasing the marginal tax rate on a tax bracket increases the tax burden for all agents above that bracket; as we assume no income effects, increasing the tax rate near the bottom mechanically increases tax revenue from nearly all agents without a welfare loss at the bottom, which is beneficial in the Rawlsian social welfare function.

Figure I

Optimal Marginal Income Tax Schedules with Inequality Externalities: Rawlsian



Optimal Marginal Income Tax Schedules with Inequality Externalities: Semi-Utilitarian



Notes: Optimal marginal tax rates for various inequality externality magnitudes η, where inequality is (a) a negative externality (social bad), (b) a positive externality (social good). The social planner is Rawlsian (above) and semi-Utilitarian (below) with $W = \beta e^{-U/\beta}$ and $\beta = 10$. The two cases converge when moving towards the optimal top tax rates. The productivity distribution is the 2018 U.S. hourly wage distribution. The wage distribution is calculated from the 2018 Annual Social and Economic Supplement of the Current Population Survey. Inequality aversion estimates indicate $\eta = 1.0$. The solid line, $\eta = 0$, is the standard case of no inequality externality. Further explanation of η is in Table II. The elasticity of labor E_L is 0.3.

in the negative externality case. This is due to these diminished revenue effects at the bottom in the semi-Utilitarian case. As the revenue benefits are lower, the mechanical equality benefits from tax raises is large enough to visibly increase tax rates.

We note that, due to the granularity of the empirical wage distribution, we do not properly observe the Pigouvian internalization at the bottom (which overtakes the mechanical effect locally). Bottom tax rates are in fact likely to decrease with a negative externality. In the lognormal case with the semi-Utilitarian SWF, which is not shown but approximates the empirical wage distribution at the bottom with a higher granularity, a negative externality of $\eta = 2.0$ decreases the bottom tax rate from 53% to 29%.

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

III.B 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, equality concerns have a large effect on the optimal top tax rate.

Equality concerns and revenue considerations differ in their impact 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. In other social welfare functions the marginal welfare of income is usually small and unchanging in the top half of the distribution.

Equality concerns are naturally different: *where* the income is taken from is of key importance. The policy effects of these equality concerns generally increase as one approaches the top of the distribution. The tax consequence of internalizing the externality (proportional to $\kappa(n) = 2F(n) - 1$ in our specification) reaches a maximum at the top. The tax consequence of the increased taste for equality (proportional to $\bar{\kappa}(n) = F(n)$ in the absolute Gini case) also reaches a local maximum at the top if the Pareto parameter is stable.[33] These observations imply that top taxation is particularly affected by the externality, which is consistent with our numerical findings.

This is clearly true when the top tax rate is zero in the standard model, which is true with a bounded wage (or income) distribution. It is also true, however, with the more commonly used assumption of an unbounded wage or income distribution, which leads to non-zero optimal marginal top tax rates even in the standard case.

To illustrate the point, consider the consequences of higher taxation in the upper part of the income distribution. The revenue benefits here are almost constant as we move upwards; one dollar of revenue remains one dollar of revenue regardless of who pays for it, and assuming no large changes in social weights, the main question normally becomes one of revenue-maximization. However, equality effects per dollar increase over the distribution. The agents targeted are increasingly only high-income agents, who have an increasingly large impact on inequality metrics.

[33] A relatively mild condition towards the top, which holds in empirical data.

Equality considerations are proportionally more important at the top than at the middle (where equality considerations are smaller) or at the bottom (where revenue considerations are large and equality considerations are small due to the imperfect targeting of bottom externalities). This is despite using the Gini coefficient as the inequality metric. The Gini is often considered as over-weighting middle-income inequalities, which dampens our effects at the top without a large impact elsewhere (discussed in Appendix E.II). When we use other inequality metrics, such as those in the S-Gini family or a family approximating top income shares, the effects of the externality become even more localized at the top.

This implies 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.A 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 wage, 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%, 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%.[34]

As we show in Tables III and IV, varying the value of the inequality-sensitivity parameter η has a large effect on the top optimal income tax rates. This is particularly true with a positive inequality externality (when inequality is a social good). These changes are generally larger than the effects from changing ϵ and $1/\alpha$. 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 wage"). 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

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

Table III

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	0.50	68	69	70	71	73	74	76	77	79	81	83	85
to	0.75	73	73	74	76	77	78	79	80	82	84	85	87
inequality	1.00	76	77	78	79	80	81	82	83	84	86	87	89
η	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 7 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 IV

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	0.50	67	68	69	70	71	73	76	79	83	90
to	0.75	71	72	73	74	76	77	79	82	86	91
inequality	1.00	75	76	76	77	79	80	82	84	88	92
η	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 7 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. The elasticity of labor E_L is 0.3 in the remainder of the paper. The standard no-externality case is in bold.

rates at the top quickly fall below 50% and begin approaching zero. We call this the Rawlsian conservative; the argument that a low top tax rate will lead to the highest possible utility for the worst-off agent.

Both of these intuitive arguments are sometimes 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 from standard results. The Rawlsian numerical simulations in Section III.A 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 if, using the framework in Equation 5,[35]

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

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 9 is zero above the very bottom earner. Thus, using the Gini values, the inequality simplifies to

[35]In the most general framework, see Appendix D, this is equal to,

$$\gamma\left[\kappa(n) + \frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty \left[\frac{\kappa(p)}{u_{x(p)}}\right] f(p)dp\right] > \frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty \left[W'(U(p))\right] f(p)dp. \quad (9)$$

$$\frac{F(z)}{\alpha(z)\epsilon(z)} > 1 - 2F(z), \tag{10}$$

which is independent of η and holds for any income above the median.[36]

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

Standard models of tax design implicitly assume that inequality has no societal effects. Remedying this issue requires an externality term in the individual utility function rather than an adjustment to the social welfare function. Introducing such an externality component has important modeling implications. Most importantly, *equality itself* becomes a policy goal. We present three new insights to the optimal income taxation literature, all of which are relevant for real-world tax design.

First: Optimal tax rates depend on the magnitude of the inequality externality, particularly at the top of the distribution. Top tax rates are largely determined by the magnitude of the inequality externality. We find theoretical support for many common policy arguments previously unsupported by economic theory; 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. This finding implies that the inequality externality is an important factor for public policy, and a potential source of political disagreement.

Second: The optimal tax structure becomes unambiguously more progressive with the introduction of a negative inequality externality. This increased progressivity is a general consequence of the inequality externality and comes from two sources: (i) the internalization of the individual externality on consumption, which is positive for low earners and negative for top earners, and (ii) society's increased willingness to pay for redistribution, which increases tax progressivity even if the agents make the right work decision. 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. The reverse applies for a positive externality.

Third: Theoretical results change substantially when introducing an inequality externality. Optimal rates above the revenue-maximizing Laffer rate are observed, particularly at the top. This is driven by negative social welfare from top incomes, which can be justified in the context of distorting inequalities. 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

[36] 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.

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 Quotes: Inequality as an externality

Plato (est. 360 B.C.): "In a state which is desirous of being saved from the greatest of all plagues [...] here should exist among the citizens neither extreme poverty, nor, again, excess of wealth, for both are productive of both these evils."

Aristotle (est. 350 B.C.): "It is clear then that those states in which the middle element is large, and stronger if possible than the other two (wealthy and poor) together, or at any rate stronger than either of them alone, have every chance of having a well-run constitution."

de Tocqueville (1835): "Men living at such times [of equality] have a natural bias to free institutions."

Keynes (1919): "In fact, it was precisely the inequality of the distribution of wealth which made possible those vast accumulations of fixed wealth and of capital improvements which distinguished that age from all others"

Roosevelt (1939): "For too many of us the political equality we once had won was meaningless in the face of economic inequality. [...] For too many of us life was no longer free; liberty no longer real; men could no longer follow the pursuit of happiness."

Sen (2001): "I believe that virtually all the problems in the world come from inequality of one kind or another."

OECD (2006): "Inequalities may create incentives for people to improve their situation through work, innovation or acquiring new skills."

Obama (2011): "This kind of inequality – a level that we haven't seen since the Great Depression – hurts us all."

Johnson (2013): "Indeed some measure of inequality is essential for the spirit of envy and keeping up with the Joneses that is, like greed, a valuable spur to economic activity."

Pope Francis (2014): "Inequality is the root of social evil."

Sanders (2014): "A nation will not survive morally or economically when so few have so much, while so many have so little."

Lagarde (2014): "The principles of solidarity and reciprocity that bind societies together are more likely to erode in excessively unequal societies."

Greenspan (2014): "You can see the deteriorating impact of [inequality] on our current political system."

Draghi (2017): "Is [inequality] a seriously destabilising factor that we should cope with? Yes it is."

Wolf (2019): "Inequality makes politics far more fractious, undermines social mobility; weakens aggregate demand and slows economic growth."

B 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 $W'(U(n))$ 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.[37]

(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.

[37] 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, 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*, see Section II.C.

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 also present a proof below to show that appropriate social weights cannot supplant an inequality externality.

B.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 $g_i \neq 0$:

$$\frac{\partial U(x_i, h_i, \theta(\boldsymbol{x}))}{\partial h_i} = \frac{\partial U(x_i, h_i, \theta(\boldsymbol{x}))}{\partial x_i} + \frac{1}{g_i} \int_j g_j \frac{\partial U(x_j, h_j, \theta(\boldsymbol{x}))}{\partial \theta(\boldsymbol{x})} \frac{\partial \theta(\boldsymbol{x})}{\partial x_i} dj \tag{11}$$

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 11 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 11 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.

C 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).[38] 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 $d\bar{\theta}_M = -\overline{\kappa}(z)\,[1-F(z)]\,dz\partial\tau$.[39]

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 income by an amount $\partial z = -\epsilon(z)z\partial\tau/\,(1-\tau(z))$. $\epsilon(z)$ 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 $-dz\partial\tau\cdot\epsilon(z)zf(z)\tau(z)/\,(1-\tau(z))$. 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.[40] The behavioral response thus has an effect on the post-tax income inequality metric as $d\bar{\theta}_B = -\kappa(z)\cdot dz\partial\tau\cdot\epsilon(z)zf(z)$, as the weight of each of these $f(z)dz$ individuals in the income inequality metric is $\kappa(z)$.

The total revenue effects are:

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

The direct welfare effect through the individual income channels is $\int_j g_j dRdj$ 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 $dM+dB+dW = dR\cdot\int_j g_j dj - dz\partial\tau\int_{\{j:z_j\geq z\}} g_j dj$.

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

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

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 $dI = \int_j g_j\cdot(-\eta)\cdot\partial\bar{\theta}\cdot dj = -\eta\cdot\partial\bar{\theta}\cdot\int_j g_j dj$.

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

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

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

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:

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

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 5.

D Analytical Solution of the OIT Problem

We write individual utility as;

$$U(x, h, \bar{\theta}) = u(x) - V(h) - \Gamma(\bar{\theta}) \tag{12}$$

where u is the utility of consumption (after-tax income), V is the disutility of work and Γ is the disutility of inequality. Equation (12) 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

$$x(n), h(n). \tag{13}$$

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,

$$\max_{T(\cdot)} \int_{\underline{n}}^{\overline{n}} W(U(x(n), h(n), \bar{\theta}))dF(n). \tag{14}$$

Notice that formulating individual utility as (12) 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 (14) 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:

$$R \leq \int_{\underline{n}}^{\overline{n}} T(nh)f(n)dn. \tag{15}$$

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:

$$u(x(n)) - V(h(n)) \geq u(x(\hat{n})) - V(h(\hat{n})) \tag{16}$$

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 (16) 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 (13). But it is also partly the result of decisions by the social planner, captured in the tax function T and therefore embedded in (13). We can represent this relationship as the following *inequality condition:*

$$\bar{\theta} = I(\boldsymbol{x}, F) \tag{17}$$

where $I(\cdot,\cdot)$ is an inequality measure, $\boldsymbol{x}(\cdot)$ is the full set of consumption choices from (13) 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:

$$I_{\text{Gini}}(\boldsymbol{x}, F) = \int_0^\infty \kappa(n)x(n)dF(n), \tag{18}$$

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

$$\kappa(n) = 2F(n) - 1 \tag{19}$$

is an expression for the weight of the agent in the Gini.[41] Expression (18) shows that the absolute

[41] 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

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 (18), condition (17) becomes

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

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,

$$\frac{dU}{dh} = 0 = u'c'n - V', \tag{20}$$

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

$$\frac{dU}{dn} = u'c'h \tag{21}$$

Taken together these two imply :

$$\frac{dU}{dn} = \frac{V'h}{n} =: g\,(n) \tag{22}$$

We can write $T = nh - x$, where x is after-tax consumption.[42] 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 $\bar{\theta}$.

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 D.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.

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

The Lagrangian of the full problem classified in Equations 14-12 is,

$$L = \int_{\underline{n}}^{\overline{n}} W(U(n))f(n)dn + \lambda(\int_{\underline{n}}^{\overline{n}} [nh(n) - x] f(n)dn)$$
$$+ \int_{\underline{n}}^{\overline{n}} \alpha(n) \left[\frac{dU}{dn} - g(n)\right] dn + \gamma \left[\bar{\theta} - I_{Gini}\right] \tag{23}$$

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:

$$L = \int_0^{\infty} [W(U(n)) + \lambda(nh(n) - x)] f(n) - \alpha(n)g(n) - \alpha'(n)U(n)dn$$
$$+\alpha(\infty)U(\infty) - \alpha(0)U(0) + \gamma \left[\bar{\theta} - I_{Gini}\right] \tag{24}$$

We introduce the Gini coefficient in the form,

$$I_{Gini} = \int_0^{\infty} [2F(n) - 1] xf(n)dn = \int_0^{\infty} \kappa(n)xf(n)dn \tag{25}$$

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:

$$L = \int_0^{\infty} \Bigg[(W(U(n)) + \lambda [nh(n) - x] - \gamma\kappa(n)x) f(n) - \alpha(n)g(n)$$
$$-\alpha'(n)U(n) \Bigg] dn + \alpha(\infty)U(\infty) - \alpha(0)U(0) + \gamma\bar{\theta} \tag{26}$$

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.[43] 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 $y_{\bar{\theta}}$.[44] The first order conditions are the following:

$$U: \quad 0 = \left[W'(U(n)) - \lambda y_U\right] f(n) - \alpha'(n) - \gamma\kappa(n)f(n)y_U \tag{27}$$

$$h: \quad 0 = \lambda(n - y_h)f(n) - \alpha(n)\frac{V_{hh}h + V_h}{n} - \gamma\kappa(n)f(n)y_h \tag{28}$$

$$\bar{\theta}: \quad 0 = \gamma - \int_0^{\infty} \gamma\kappa(n)f(n)y_{\bar{\theta}}dn - \int_0^{\infty} \lambda y_{\bar{\theta}}f(n)dn \tag{29}$$

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

[44] 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 $y_U = \frac{1}{u_x}$.

In the FOC for h we have used that $g = \frac{V_h h}{n}$ from Equation (22), and that $\frac{dg}{dh} = \frac{V_{hh}h + V_h}{n}$. Equation 29 implies,

$$\frac{\gamma}{\lambda} = \frac{\int_0^\infty \frac{\Gamma_{\bar{\theta}}}{u_x} f(n) dn}{1 - \int_0^\infty \frac{\Gamma_{\bar{\theta}}}{u_x} \kappa(n) f(n) dn} \tag{30}$$

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 27:[45]

$$\alpha(n) = \int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p)) \right] f(p) dp \tag{31}$$

And substitute this into Equation (28):

$$0 = \lambda(n - y_h) f(n) - \gamma\kappa(n) f(n) y_h - \frac{V_{hh}h + V_h}{n} \int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p)) \right] f(p) dp \tag{32}$$

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

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:

$$\frac{t}{1-t} = \frac{\gamma}{\lambda}\kappa(n) + \frac{\zeta_n u_{x(n)}}{\lambda f(n) n} \int_n^\infty \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W'(U(p)) \right] dF(p). \tag{34}$$

Here $\frac{\gamma}{\lambda}$ is the price of inequality in terms of public funds (see Equation 30). If inequality is a negative externality (a public bad), γ will generally be large and positive.[46] 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,[47] 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

[45] 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.

[46] If we assume a linear inequality externality of the form $\Gamma(\theta) = \eta\theta$ then $\frac{\gamma}{\lambda} = \eta$ (see Equation 30).

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

externality.

$$\frac{t}{1-t} = \frac{\gamma}{\lambda}\left[\kappa(n) + \frac{\zeta}{f(n)n}\int_n^\infty \frac{u_{x(n)}}{u_{x(p)}}\kappa(p)dF(p)\right] + \frac{t_i}{1-t_i} \tag{35}$$

For clarity let us assume a linear homogeneous inequality externality ($\Gamma(\bar{\theta}) = \eta\bar{\theta}$) and quasi-linearity in consumption.[48] The optimal tax rate condition simplifies to:

$$\frac{t}{1-t} = \eta\kappa(n) + \eta\left(1 + \frac{1}{E_L}\right)\Pi(n)F(n) + \frac{t_i}{1-t_i}, \tag{36}$$

where we denote the distributional thinness measure $\frac{1-F(n)}{f(n)n}$ as $\Pi(n)$.[49] This formula is functionally equivalent to the form we found with the small-perturbations method (Equation 5), but uses the exogenous wage n instead of the endogenous earnings z.

D.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 $\Omega(n) = x(n) - V(\frac{z(n)}{n})$. Here $\Omega(n) \geq \Omega(\hat{n})\,\forall\, n, \hat{n}$ is equivalent to the IC constraint. The problem becomes

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

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

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

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

[48] 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 (35) simplifies as $\int_n^\infty (2F(n) - 1)\, dF(n) = F(n) - F(n)^2$.

[49] 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; $\varepsilon_{P,n} = \frac{n}{1-F(n)}\frac{d(1-F)}{dn} = -\frac{nf(n)}{1-F(n)}$.

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.[50] 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.

E Additional notes for Section

E.I. Theoretical ability distributions

We present Rawlsian optimal marginal income tax rates from two theoretical skill distributions in Figure II. The first is is a Pareto distribution with $\alpha(n) = 2.0$, which becomes identical to the empirical case at the top of the distribution.[51] The second 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 IIa) illustrates the effect of equality considerations 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 compared to the no-externality case. The goal of this tax subsidy is to make individuals internalize that their increased labour supply leads to positive societal outcomes. This effect is present in all the simulations, but is too small to be detected when using other skill distributions. The effect leads to negative income tax rates at the bottom under certain specifications.

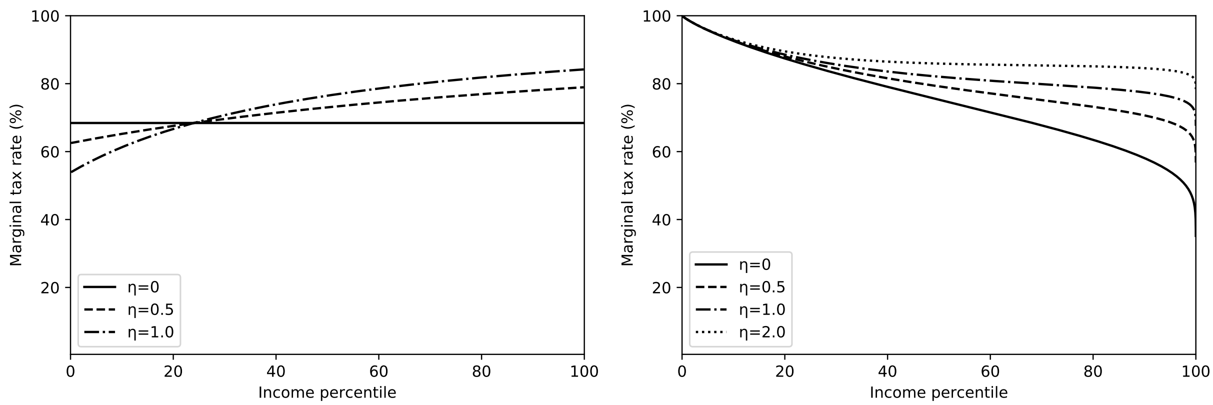
The lognormal case further illustrates the localized effects at the top of the distribution. The standard top marginal tax rate, in the lognormal case, is 0%. With an inequality externality of $\eta = 2.0$ that increases to 67%. This illustrates the Pigouvian correction at the top, and is salient given 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% to 79%.

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

[51] Under this Pareto distribution, second-order conditions fail for $\eta = 2.0$. This is therefore not plotted.

Figure II

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 $a = 2$, (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 II 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.

E.II. Varying inequality metrics

In the main text we used the absolute Gini coefficient for our measure of inequality. Here we explore two different families of inequality metrics. The first shares similarities with top income shares, whereas the second (the S-Gini) approximates the Gini with a larger focus on either end of the distribution. The distributional weights implied by both families are plotted in Figure III.[52]

1 Approximating top income shares The first family of inequality metrics has some of the properties of top income shares. It is,

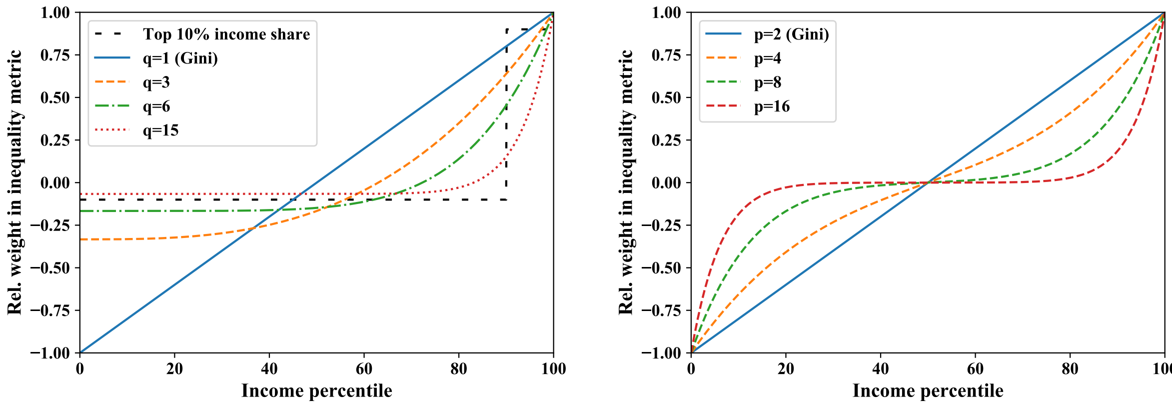
$$\bar{\theta} = \int_0^\infty \left[(q+1)F(n)^q - 1\right] x(n) dF(n), \; q \in \mathbb{N}. \tag{37}$$

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.

[52]The weights in Figure III 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 III

Weights for Families of Inequality Metrics



Note: Consumption weights for inequality metrics used in Appendix E.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 37. The top 10% income share is plotted in dotted black for reference. (b) The S-Gini family from Equation 39.

The resulting optimal tax rates with the utility function in 8 become,

$$\frac{t}{1-t} = \eta_q \left[((q+1)F(n)^q - 1) + (1 + \frac{1}{E_c}) \frac{1}{f(n)n} \left[1 - F(n)^q\right] F(n) \right] + \frac{t_{orig}}{1 - t_{orig}}. \tag{38}$$

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.[53]

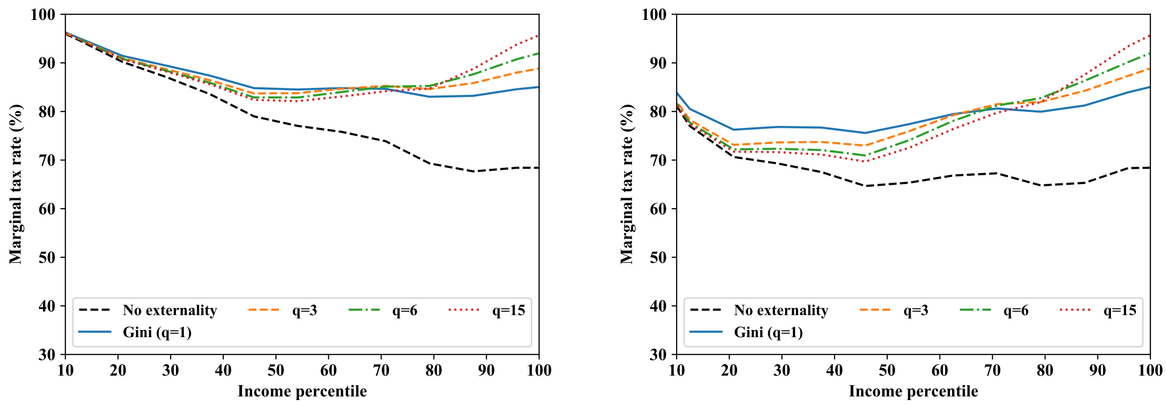
In Figure IV we show the effect of increasing the inequality metric from $q = 1$ (the Gini) to values increasingly focused on top incomes. To focus on the effects of changing the inequality metric we keep the externality value constant at the upper bound of the median inequality aversion range from Carlsson et al. (2005). The no-externality case is shown as a reference. The ability distribution is the empirical wage distribution, and the SWF is (a) Rawlsian and (b) $W = \beta e^{-U/\beta}$ with $\beta = 10$.

Increasing q leads to the effects of the externality becoming more concentrated towards the top of the distribution. This leads to the tax changes being more concentrated towards the top. This is particularly notable in the non-Rawlsian case, where the decreased effect of the externality at the bottom is clearly noticeable. This strengthens the results discussed in the main text. Other externality values and ability distributions (not shown here) lead to similar results.

[53] 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.

Figure IV

Optimal Taxation with Inequality Externalities: Top Income Share Metric Family



Note: Optimal marginal tax rates for various negative inequality metrics of the family such that $\int_0^\infty [(q+1)F(n)^q - 1]\, x(n)dF(n)$, $q \in \mathbb{N}$. The social planner is (a) Rawlsian, and (b) $W = \beta e^{-U/\beta}$ with $\beta = 10$. The magnitude of the inequality externality is held constant for all q at the upper bound of the median value from the empirical inequality aversion estimates in Carlsson et al. (2005). The productivity distribution is the empirical wage distribution from the main text. See Figure III for an explanation of the inequality metrics magnitudes. In particular, larger q indicates that top incomes are increasingly weighted. The black dotted line is the standard case of no inequality externality. The elasticity of labor E_L is 0.3.

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 V. 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 more intuitive impacts on $\frac{t}{1-t}$. All social planners are Rawlsian.[54]

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.

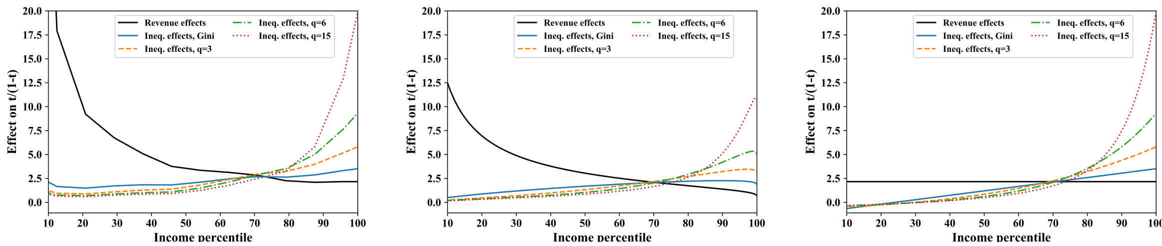
$$\bar{\theta} = \int_0^\infty [F(n)^p - (1-F(n))^p]\, x(n)dF(n),\ p \geq 2. \tag{39}$$

When $p = 2$, this becomes the absolute Gini coefficient. This family also retains the beneficial

[54]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.

Figure V

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



Note: Effects on $\frac{t}{1-t}$ for various negative inequality metrics $\int_0^\infty \left[(q+1)F(n)^q - 1\right] x(n)dF(n),\ q \in \mathbb{N}$. 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 III 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.

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 8 are,

$$\frac{t}{1-t} = \eta_p \left[(F(n)^p - (1-F(n))^p) + (1 + \frac{1}{E_c}) \frac{1}{f(n)n} \nu \right] + \frac{t_{orig}}{1 - t_{orig}}, \tag{40}$$

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

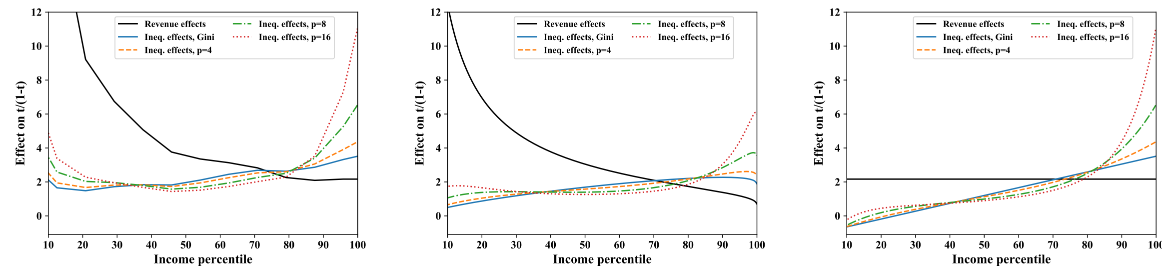
In Figure VI we show the effect of changing p on $\frac{t}{1-t}$ with the same methodology as in Figure V. Increasing p again leads to larger effects towards the top of the distribution and relatively small changes at the bottom. It is notable that the effects at the bottom remain small despite the increased magnitude of the positive externality on these individuals' income. This is driven by two factors. First, bottom incomes are smaller than those at the top by definition and thus have a smaller effect on inequality metrics, even when both are weighted symmetrically. Second, every individual is affected by marginal tax rate changes at the bottom, whereas marginal tax rate changes at the top impact only those at the top. Thus 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.

Three of the four points noted in the previous subsection also hold for the S-Gini. The one exception is the sign of the change in the bottom half when changing the metric, which is ambiguous due to the counteracting equality effects.

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

Figure VI

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 III 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 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.

E.III. A squared inequality externality function $\Gamma(\theta)$

Our framework is sufficiently general for other functional forms of $\Gamma(\bar{\theta})$. Let us use $\Gamma(\bar{\theta}) = \eta(\bar{\theta} - \theta_{opt})^2$, such that:

$$U(x, h, \bar{\theta}) = x - \frac{h^{\left(1+\frac{1}{E_c}\right)}}{\left(1+\frac{1}{E_c}\right)} - \eta(\bar{\theta} - \theta_{opt})^2 \tag{41}$$

The resulting optimal tax rates are:

$$\frac{t}{1-t} = 2\eta(\bar{\theta} - \theta_{opt}) \left[\kappa(n) + \frac{\zeta}{f(n)n} \int_n^\infty \kappa(p) f(p) dp\right] + \frac{t_{orig}}{1 - t_{orig}} \tag{42}$$

Comparing these tax rates to Equation 35, we see that the effect of the inequality externality is attenuated by a factor of $2(\bar{\theta} - \theta_{opt})$. 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.

Also note that this solution is endogenous, as $\bar{\theta}$ depends on the tax schedule. We 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.

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.

Aristotle (1981). *The Politics, Book IV*. Penguin UK. Estimated written 350 B.C.. Translated by Trevor J. Saunders.

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.

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.

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.

de Tocqueville, A. (1835). *Democracy in America*. New American Library. Edited by Richard D. Heffner, 1984.

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.

Draghi, M. (2017). Europe's inequality highly destabilising, ECB's Draghi says. Lisbon public comments, https://www.reuters.com/article/ecb-policy-draghi/europes-inequality-highly-destabilising-ecbs-draghi-says-idUSL3N1JN44D.

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.

Greenspan, A. (2014). Comments: National Association of Business Economists conference. https://www.thefiscaltimes.com/Articles/2014/02/25/These-Top-Economists-All-Agree-Biggest-Problem-US-Faces.

Johnson, B. (2013). Boris Johnson invokes Thatcher spirit with greed is good speech. Comments at 2013 Margaret Thatcher lecture.

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.

Keynes, J. M. (1919). *The economic consequences of the peace*, volume 178. Springer.

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

Lagarde, C. (2014). Economic inclusion and financial integrity - an address to the conference on inclusive capitalism. *London, UK. Retrieved April*, 19:2016.

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.

Obama, B. (2011). Comments: Osawatomie High School. https://www.theatlantic.com/politics/archive/2011/12/obamas-income-inequality-speech/334664/.

OECD (2006). *OECD Factbook 2006: Economic, Environmental and Social Statistics.* OECD Publications.

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.

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

Pope Francis (2014). @Pontifex Twitter. https://twitter.com/pontifex/status/460697074585980928.

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

Roosevelt, F. D. (1939). Acceptance speech at the Democratic National Convention.

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.

Sanders, B. (2014). @SenSanders Twitter. https://twitter.com/sensanders/status/426740006905200640.

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.

Sen, A. (2001). An interview with Amartya Sen. Interview by David Barsamian.

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*.

Wolf, M. (2019). How to reform today's rigged capitalism. *Financial Times*. December 3 2019.

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