The Inequality Deflator:
Interpersonal Comparisons without a Social Welfare Function

Nathaniel Hendren*

July, 2014

Abstract

This paper develops a tractable method for resolving the equity-efficiency tradeoff that modifies the Kaldor-Hicks compensation principle to account for the distortionary cost of redistribution. Weighting measures of individual surplus by the inequality deflator corresponds to searching for local Pareto improvements by making transfers through the income tax schedule. Empirical evidence consistently suggests redistribution from rich to poor is more costly than redistribution from poor to rich. As a result, the inequality deflator weights surplus accruing to the poor more so than to the rich. Regardless of one's own social preferences, surplus to the poor can hypothetically be turned into more surplus to everyone through reductions in distortionary taxation. I estimate the deflator using existing estimates of the response to taxation, combined with a new estimation of the joint distribution of taxable income and marginal tax rates. I show adjusting for increased income inequality lowers the rate of U.S. economic growth since 1980 by roughly 15-20%, implying a social cost of increased income inequality in the U.S. of roughly $400 billion. Adjusting for differences in income inequality across countries, the U.S. is poorer than countries like Austria and the Netherlands, despite having higher national income per capita. I conclude by providing an empirical framework for characterizing the existence of local Pareto improvements from government policy changes.

1 Introduction

The measurement of societal well-being is an old endeavor in economics. While the canonical utility-maximizing framework provides a fairly straightforward, if controversial, method for measuring individual well-being, aggregating across individuals is notoriously more difficult.

*Harvard University, nhendren@fas.harvard.edu. I am deeply indebted to conversations with Louis Kaplow for the inspiration behind this paper, and to Sarah Abraham, Alex Bell, Alex Olssen, and Evan Storms for excellent research assistance. I also thank Daron Acemoglu, Raj Chetty, Amy Finkelstein, Ben Lockwood, Henrik Kleven, Patrick Kline, Jim Poterba, Emmanuel Saez, Matthew Weinzierl, Glen Weyl, Ivan Werning, and Floris Zoutman, along with seminar participants at Berkeley, Harvard, MIT, Michigan, and Stanford for very helpful comments. The opinions expressed in this paper are those of the author alone and do not necessarily reflect the views of the Internal Revenue Service or the U.S. Treasury Department. This work is a component of a larger project examining the effects of tax expenditures on the budget deficit and economic activity, and this paper in particular provides a general characterization of the welfare impact of changes in tax expenditures relative to changes in tax rates (illustrated in Section 6). The empirical results derived from tax data that are reported in this paper are drawn from the SOI Working Paper "The Economic Impacts of Tax Expenditures: Evidence from Spatial Variation across the U.S.", approved under IRS contract TIRNO-12-P-00374.

Aggregation is unavoidable for many normative questions: Is free trade good? What are the welfare consequences of skill-biased technological change or other forces increasing income inequality in the U.S.? How should one weight producer and consumer surplus? Interpersonal comparisons are ubiquitous; yet there is no well-agreed upon method for their resolution.

Beginning with Kaldor (1939) and Hicks (1939, 1940), a common approach is to separate issues of distribution (equity) from the sum of income or welfare (efficiency). They propose a compensation principle that led to aggregate surplus, or efficiency, as a normative criteria: if one environment delivered greater total surplus relative to the status quo, then the winners could compensate the losers through a hypothetical redistribution of income. So, comparing alternative environments required only summing up individual willingness to pay using expenditure functions. While careful attention was paid to measuring aggregate purchasing power or welfare appropriately (e.g. using price deflators), one could ignore changes in the distribution of welfare within the economy.

The focus on aggregate surplus resolves interpersonal comparisons by valuing money equally to rich and poor (Boadway (1974); Fleurbaey (2009)). Given preferences for equity, the common alternative is to use a social welfare function.[1] But, this requires the economist to specify a subjective preference for equity in order to measure social welfare. So, it is difficult to make policy recommendations based on this approach that command universal acceptance.[2]

This paper returns to the Kaldor-Hicks criteria, but adds the modification that the transfers be feasible. In particular, I require the transfers conform with Mirrlees (1971)'s observation that information constraints prevent individual-specific lump-sum taxation. It is well-known that Kaldor and Hicks envisioned researchers accounting for the distortionary impact of transfers[3], and many subsequent papers have explored the implications of doing so. For example, Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2008) modify Kaldor (1939) so that compensating transfers occur through the income tax schedule; Coate (2000) modifies Hicks (1940) by making comparisons to a feasible set of alternatives (that exclude individual specific lump-sum transfers). However, as noted by Coate (2000), what is missing from this literature is a simple shadow-price method for accounting for these distortions.[4]

[1] See Bergson (1938); Samuelson (1947); Diamond and Mirrlees (1971); Mirrlees (1976); Dreze and Stern (1987); Slemrod and Yitzhaki (2001). An alternative proposal, closer in spirit to the present paper, is to focus on characterizing the Pareto frontier for changes in the nonlinear income tax schedule (Werning (2007)). While in general it is difficult to find Pareto improvements through modifications only to the income tax schedule, the argument of this paper is that the resulting shadow prices from this problem are useful to search for Pareto comparisons for other settings.

[2] In his classic rebuff of distributional weighting, Harberger (1971) writes (p787): *"Hypothetically, one might contemplate a national income measure incorporating "distributional weights," but two obstacles stand in its way: first, the impossibility of achieving a consensus with regard to the weights, and second, the fact that most of the data from which the national accounts are built are aggregates in the first place, and do not distinguish the individuals or groups whose dollars they represent."* Fortunately, the rise of data availability has largely removed the second problem; this paper attempts to make progress on the first.

[3] Hicks writes, *"If, as will often happen, the best methods of compensation feasible involve some loss in productive efficiency, this loss will have to be taken into account"* (Hicks (1939), p712)

[4] Coate (2000) writes: *"One attraction of the social welfare function approach is that it is possible to define a set of shadow prices and instruct government agencies to implement any project making positive profits at those prices (see Dreze and Stern (1987) for the details). It is not obvious that this is possible under the efficiency approach, since implementing it requires more than local information about feasible policies and their consequences. An interesting problem for further research would be to investigate whether the efficiency approach might be approximately decentralised via a system of*

As a first step in this direction, this paper develops and estimates a shadow-price method that locally characterizes the existence of potential Pareto improvements when transfers occur through modifications to the nonlinear income tax schedule. I show one can search for these Pareto improvements by weighting standard measures of individual surplus (e.g. compensating and equivalent variation) by an inequality deflator, $g(y)$, defined at each income level, y.[5] If \$1 of surplus falls in the hands of someone earning \$$y$, this can be turned into \$$g(y)/\mathrm{n}$ of surplus to everyone in the economy (where n is the number of people in the economy). Weighting surplus by the inequality deflator constructs a hypothetical experiment whereby the surplus is redistributed equally across the income distribution using modifications to the income tax schedule.

The inequality deflator differs from unity because behavioral responses affect the government budget through fiscal externalities. In particular, empirical evidence consistently suggests that it is more costly to redistribute from rich to poor than from poor to rich. For example, Saez et al. (2012) suggest a \$1 mechanical decrease in tax liability for those facing the top marginal income tax rate has a fiscal cost of \$0.50 - \$0.75 because reducing tax rates would increase taxable earnings. At the other end of the income distribution, Hendren (2013) draws on the summaries in Hotz and Scholz (2003) and Chetty et al. (2013) and calculates that expansions of the earned income tax credit (EITC) to low earners has a fiscal cost of around \$1.14 because of behavioral responses. Hence, a dollar of surplus to the rich can be translated into \$0.44-\$0.66 to the poor. Conversely, a dollar of surplus to the poor can be translated into \$1.52-\$2.28 to the rich through a reduction in marginal tax rates and EITC distortions. Therefore, surplus to the poor should be valued roughly twice as much as surplus to the rich. The Kaldor-Hicks logic justifies this weighting regardless of one's own social preference: even if one only valued surplus to the rich, \$1 of surplus accruing to the poor can be turned into more than \$1 to the rich through modifications to the tax schedule.

Although weighting surplus by the inequality deflator corresponds to searching for local Pareto improvements, it is related to the social welfare function approach. The inequality deflator equals the average social marginal utilities of income at each income level that rationalize the status quo tax schedule as optimal. If one assumes the government is optimizing the tax schedule using these weights, the inequality deflator is the solution to the "inverse optimum" program of optimal taxation (Dreze and Stern (1987); Blundell et al. (2009); Bargain et al. (2011); Bourguignon and Spadaro (2012); Zoutman et al. (2013a,b); Lockwood and Weinzierl (2014)).[6] Intuitively, if one's own social preferences are

shadow prices which convey the cost of redistributing between different types of citizens." The present paper provides this decentralization for the case when the comparison set of policies are local changes to the nonlinear income tax schedule and illustrates how one can use these shadow prices can be applied even in settings with multi-dimensional heterogeneity (see Proposition 4).

[5]The focus on the tax modifications is motivated by the Atkinson and Stiglitz (1976) idea that in many cases this is the most efficient method for accomplishing redistribution. But, I also discuss extensions to other incentive feasible transfers. The focus on a first-order characterization circumvents the intransitivity issues that arise to second-order in the Kaldor-Hicks setup (Scitovsky (1941); Boadway (1974)).

[6]The approach is also related to the large literature on distributional weighting of the marginal cost of public funds (MCPF). Instead of starting with a social welfare function, this paper starts with the Kaldor Hicks principles and never relies on an assumption that the government is choosing policy to optimize its objectives. But, it turns out that the results in this paper suggest that the literature estimating the *heterogeneity* in the MCPF of non-budget neutral policies across the income distribution can be used as an inequality deflator, provided one adopts the non-budget neutral definition of

willing to pay more than (less than) \$2 to the rich to transfer \$1 to the poor, then one might prefer a more (less) redistribution through the tax schedule. But, regardless of one's own opinion (or the policymaker's opinion) about whether society should give more money to the poor, the Kaldor-Hicks logic justifies using these weights as if it were the relevant social welfare function.

I derive the inequality deflator in a general setting with multi-dimensional heterogeneity. In doing so, I show that the same inequality deflator applies regardless of whether income inequality is the result of differences in preferences or differences in abilities (or both). But, a difficult set of issues arise when two different people have the same income but different surplus. It is infeasible to use the income tax schedule to provide different sized transfers to those with the same income.[7] In such cases, the inequality deflator can be used to characterize the existence of local Pareto improvement, but it may not be feasible to provide a local Pareto ranking. I offer several potential paths forward, such as extending the transfers to multiple policy dimensions, discussed further in Section 3.7.

To provide a precise estimate of the inequality deflator at each income level, I provide a new representation of the fiscal externality associated with marginal changes to tax policies.[8] The marginal cost of taxation depends on the joint distribution of marginal tax rates, local Pareto parameter of the income distribution (i.e. the shape of the income distribution), and taxable income elasticities. In the presence of multi-dimensional heterogeneity, I show this fiscal externality depends on population-average taxable income elasticities conditional on income, consistent with an intuition provided in Saez (2001).

I provide an estimate of the inequality deflator by taking estimates of taxable income elasticities from existing literature, combined with a new estimation of the joint distribution of marginal tax rates and the income distribution using the universe of U.S. income tax returns in 2012. The use of population tax records allows me to observe each filer's marginal tax rate and then non-parametrically estimate the shape of the income distribution conditional on each marginal tax rate, which is a key input into the formula for the inequality deflator. I choose estimates of the taxable income elasticity consistent with existing literature – discussed in detail in Section 4. Existing estimates suggest the taxable income elasticity is perhaps relatively stable across the income distribution and across tax reforms (Chetty (2012)). But, the shape of the income distribution and size of the marginal tax rates vary considerably across the income distribution, leading to heterogeneous fiscal externalities from

the MCPF put forth in Kleven and Kreiner (2006) and used in Hendren (2013). In contrast, the traditional definition of the MCPF (a.k.a the Atkinson-Stern-Stiglitz-Dasgupta definition ; see Stiglitz and Dasgupta (1971); Atkinson and Stern (1974); Ballard and Fullerton (1992)), considers policies which spend the tax revenue on a public good received by the individual being taxed. Such estimates cannot be combined to provide a redistributive policy akin to the redistribution envisioned by Kaldor and Hicks.

[7]Interestingly, the social welfare function interpretation of the inequality deflator is sustained if and only if one is willing to assume that social marginal utilities of income are constant conditional on taxable income. Without this restriction, the inequality deflated surplus does not bound the implicit social welfare impact even if one wanted to use the implicit social welfare weights that rationalize the tax schedule as optimal (see Corollary 3). But, the inequality deflator retains its marginal cost interpretation as a tool for searching for potential Pareto improvements.

[8]The representation generalizes existing elasticity representations of the marginal cost of taxation in the presence of intensive and extensive margin responses (e.g. Bourguignon and Spadaro (2012); Zoutman et al. (2013a,b)) by allowing for essential heterogeneity in the utility function (as opposed to assuming uni-dimensional heterogeneity and the Spence-Mirrlees single crossing property).

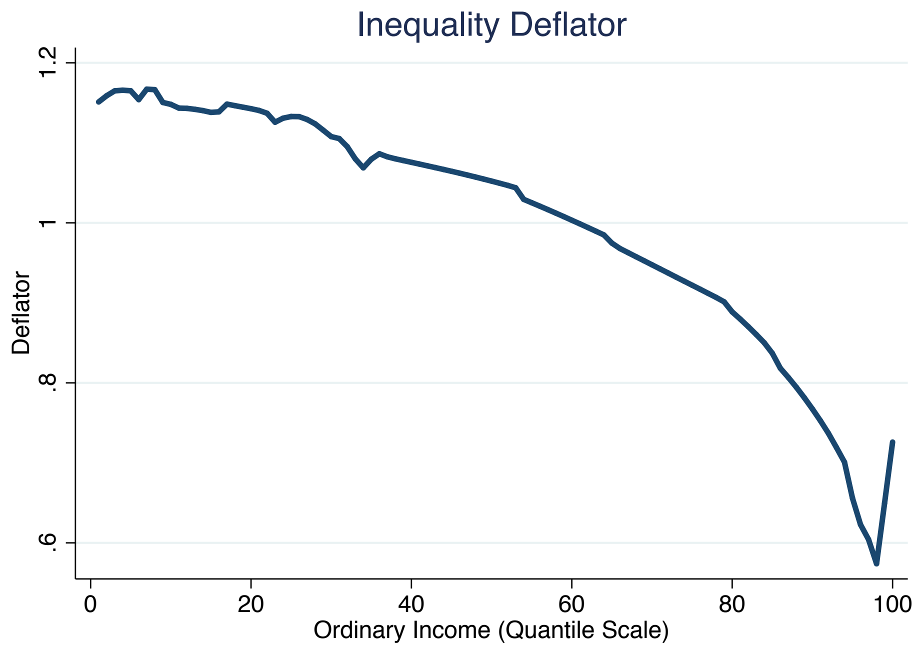

Figure 1: Inequality Deflator, $g(y)$. \$1 of surplus falling to those earning y can be turned into $g(y)/n$ surplus to everyone. Estimation of $g(y)$ discussed in Section 4.

changes in the tax schedule.

Figure 1 presents the baseline estimates of the inequality deflator. The values range from 1.15 near the bottom of the income distribution to near 0.6 in the 98th percentile of the income distribution. This means that if \$1 of surplus were to fall to the bottom of the income distribution, it can be turned into \$1.15/n of surplus to everyone. Conversely, if \$1 of surplus accrues to the 98th percentile, it can be turned into \$0.60/n to everyone through modifications to the income tax schedule. In this sense, surplus is more socially valuable if it accrues to the bottom of the income distribution.

The inequality deflator has several additional features to note. First, the fact that the deflator is everywhere positive implies there are no Laffer effects: changes to the ordinary income tax rate alone cannot generate Pareto improvements.[9] Second, the slope of the deflator is steeper in the upper half of the income distribution than the lower half. This suggests it is more costly to redistribute from high-earners to median earners than from median earners to the low-earners. Finally, the deflator declines towards the 98th percentile of the income distribution, but then exhibits a non-monotonicity at the top 1% of the income distribution. This suggests current tax rates implicitly value resources more in the top 1% (greater than ~\$350K) of the income distribution relative to the 98th percentile (~\$250K-\$350K in ordinary income).[10]

[9] This test is similar to that of Werning (2007) who searches for Pareto improvements to the existing nonlinear tax schedule.

[10] The baseline specification assumes that the elasticity of taxable income is roughly constant and equal to 0.3 above the 95th percentile. I show in Section 4.4 that monotonicity is restored if one assumes that the elasticity of taxable income increases from 0.3 at the 98th percentile to 0.5 at the 99th/100th percentile.

I illustrate how to apply the inequality deflator by using it to compare income distributions.[11] While it is common to use price deflators (e.g. CPI, PPP, etc.) to adjust income comparisons for differences in the aggregate purchasing power of an economy, the inequality deflator allows one to adjust for differences in the *distribution of individual purchasing power*. I illustrate this with two applications: historical changes within the U.S. and comparisons across countries.[12]

It is well known that the U.S. has experienced a significant increase in income inequality, especially in the top 1% (Piketty and Saez (2003)). I show that, although mean household income is roughly \$18,300 higher per household relative to 1980 (in 2012 dollars), inequality-deflated growth is only \$15,000. In other words, if the U.S. were to modify the tax schedule so that every point along the income distribution experienced equal gains, \$3K of this \$18K surplus would disappear, evaporating 15-20% of the mean household income growth.[13] On the one hand, this difference is not enormous – the estimates suggest the U.S. is still significantly "richer" today than in 1980, even after adjusting for increased inequality. But, aggregating across the roughly 120M households in the U.S., this implies a total adjustment for rising income inequality in the U.S. of \$400B. Put differently, the modified Kaldor-Hicks logic suggests that the U.S. should be willing to pay \$400B for a policy that led to the same aggregate 2012 after-tax income in the U.S. but that did not also have the increased income inequality.

It is also well known that the U.S. has greater income inequality than many other countries, especially those in western Europe, but has higher per capita income. In particular, the U.S. has roughly \$2,000 more mean household income than than Austria and the Netherlands.[14] I show that, if the U.S. were to adjust its income distribution to offer the distribution of purchasing power provided in these countries, the inequality deflator suggests it would be roughly \$227 poorer than the Netherlands and \$366 poorer than Austria.[15] In this sense, the inequality deflator provides a method for adjusting cross-country comparisons not only for differences in aggregate purchasing power, but also

[11] As a benchmark, I measure normative differences in income distributions under the assumption that they represent solely differences in purchasing opportunities, as opposed to differences in leisure; the approach could easily incorporate leisure or other utility-relevant factors given estimates of their values across the income distribution.

[12] This exercise is related to Jones and Klenow (2011) who use a parameterized utility function to measure welfare across countries and time that has log curvature in consumption. Relative to this approach, the inequality deflator allows one to make such comparisons avoiding sources of potential disagreement about the curvature of the utility function or social welfare function. For example, the 2-1 ratio of valuations of surplus to poor versus rich (shown in Figure 1) contrast with the ratios well above 10-1 that are implied by the log curvature used in Jones and Klenow (2011).

[13] Note this is a measure of the change in economic surplus (as measured by willingness to pay), not the change in GDP in response to the redistribution. The envelope theorem implies that behavioral responses to the redistribution of this \$18K surplus do not affect anyone's utility directly. But, their fiscal externalities dissolve \$3K of the \$18K surplus, so that each point of the income distribution can only be made \$15K better off relative to 1980, as opposed to \$18K better off. The measured GDP impact of redistributing \$18K surplus is likely to be larger than \$3K. But, by relying on the envelope theorem, the inequality deflator accounts for the value of leisure in response to redistribution.

[14] As a benchmark for comparison, I compare the income distributions across countries under the assumption of no differences in the value of leisure or other public goods. To the extent to which European countries offer more progressive allocation of public goods or leisure, this leads to an under-estimate of the deflation one needs to apply to make surplus comparisons between the U.S. and these countries.

[15] Similar to the U.S. comparison over time, this is a measure of individual willingness to pay, not a measure of how much lower the U.S. GDP would be relative to these countries if it were to engage in redistribution. If the U.S. were to replicate the income distribution of Austria and the Netherlands, measured GDP would be much lower than these two countries; but the inequality deflator accounts for the value of leisure in response to redistribution.

for differences in the distributions of purchasing power.

Finally, I turn to policy implications. For budget neutral policies, one can weight measures of each individual's willingness to pay for the policy change by the inequality deflator to characterize potential Pareto improvements. For non-budget neutral policy experiments, one can compare "benefits" to "costs". However, the benefits must be inequality deflated and the costs must include any fiscal externalities. This provides an empirical generalization of the "benefit principle" – taxing individuals in proportion to the benefits they receive from the policy – to cases where one does not make separability assumptions on the utility function.[16] I show that non-separability can be accounted for empirically by estimating the aggregate fiscal externality associated with the policy (so that one has an accurate measure of its total cost) and a deflated measure of benefits (so that one accounts for the unequal distribution of surplus).[17]

I illustrate this framework by asking whether the tax schedule and the EITC is a more desirable method of redistribution relative to other policies such as food stamps, housing vouchers, and job training programs. Building on the analysis of Hendren (2013), I show how one can make welfare statements about the desirability of these programs given estimates of (a) their causal effects on taxable behavior and (b) beneficiaries' willingness to pay for the government expenditure on the program.[18] For baseline estimates, the EITC appears more efficient at bringing resources to the poor than housing vouchers and food stamps. This is subject to many empirical caveats; but, these limitations are empirical, not philosophical. In the end, the inequality deflator generates welfare statements using information derived solely from individual behaviors, as opposed to social preferences of the researcher.

The rest of this paper proceeds as follows. Section 2 provides an introductory example to illustrate the main ideas. Section 3 presents the model of interpersonal comparisons and defines the inequality deflator. Section 4 discusses the estimation of the inequality deflator using the universe of U.S. income tax returns and elasticity estimates from existing literature. Section 5 applies the inequality deflator to the comparisons of income distributions. Section 6 discusses the implications for the welfare analysis of public policies. Section 7 concludes.

[16] See Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2006, 2008) for seminal work on the benefits principle. See also Kreiner and Verdelin (2012) for a summary and a comparison to the traditional MCPF approach.

[17] An empirical generalization of the benefit principle to cases where weak separability of the utility function does not hold has proven elusive. For example, Kreiner and Verdelin (2012) write *"in a general setting without separability, the optimal supply of public goods follows a modified Samuelson rule, with an additional term representing the correlation between ability and the demand for the public good – conditional on income. It is very difficult to identify this additional term empirically because correlations between demand and, respectively, ability and income are observationally equivalent but have vastly different policy implications, as first noted by Hylland and Zeckhauser (1979)."* Section 6 shows that these non-separabilities have an empirical representation in terms of the difference between the fiscal externality associated with modifying the income tax schedule and the fiscal externality associated with the public expenditure. So, one can account for them by measuring individual willingness to pay and with the cost of the policy (inclusive of these fiscal externalities).

[18] This may be different than dollar-for-dollar if, for example, individuals cannot re-trade the expenditure, as is the case with food stamps.

2 Introductory Example

To motivate the inequality deflator, suppose an alternative environment is preferred by the poor but not by the rich. Figure 2 presents the willingness to pay for this hypothetical alternative environment across the income distribution. The standard Kaldor-Hicks compensation principle would simply sum up this willingness to pay. If aggregate willingness to pay is positive, the winners could hypothetically compensate the losers from moving to the alternative environment.

But, now suppose that these transfers had to occur through modifications to the income tax schedule. Such transfers will involve distortionary costs. To illustrate this, imagine providing \$1 of a tax deduction to those with incomes in an ϵ-region near a given income level, y^*, as depicted in Figure 3. To first order, those directly affected by the transfer value them at their mechanical cost, \$1.[19] However, the cost of these transfers has two components. First, there is the mechanical cost of the transfer, \$1. But, in addition, some people will change their behavior to obtain the transfer, so that the total cost to the government will be given by $1 + FE(y)$, where $FE(y)$ is the fiscal externality resulting from the behavioral responses to the modification to the tax schedule.[20] These fiscal externalities across the income distribution will characterize the marginal cost of redistribution through the tax schedule.

Given the marginal cost of taxation, one can imagine neutralizing distributional comparisons between the status quo and alternative environments in two ways, analogous to equivalent and compensating variation. First, one can imagine that the losers have to bribe the winners in the status quo environment. This is an equivalent variation approach depicted in Figure 4. In this figure, individuals are (to first order) indifferent between the alternative environment and the modified status quo depicted by the red line. So, if the tax augmented schedule (red line) is budget feasible, one could close the resource constraint by providing a uniform benefit to everyone, as depicted in the blue line in Figure 5. Conversely, if the red line is not budget feasible, then closing the budget constraint using a uniform payment would induce a uniform cost to everyone, as depicted in Figure 6. The difference between the red and blue line will be called inequality deflated surplus. It measures how much everyone can be made better off in the alternative environment relative to the modified status quo.

In addition to the equivalent variation (EV) approach, one can also implement a compensating variation (CV) approach that modifies the tax schedule in the alternative environment. Here, the inequality deflator in the alternative environment can be used to characterize the extent to which everyone can be made better off in the modified alternative environment relative to the status quo. In applications, it may be reasonable to assume that the inequality deflator is roughly similar in the

[19] I assume those not directly affected by the transfer do not have a welfare impact from the transfer. This assumption is quite common in existing literature, but rules out potential trickle-down or trickle-up effects of taxation, along with other types of GE effects (e.g. impacts on tax wages). These effects are excluded not because they are not important, but rather because their empirical magnitudes are notoriously difficult to uncover. Extending the inequality deflator to settings with such non-localized impacts of taxation is an important direction for future work.

[20] Because of the envelope theorem, those who change their behavior to obtain the transfer will not experience a first order gain in utility. Moreover, this fiscal externality term, $FE(y)$, is not a traditional measure of marginal deadweight loss. It depends on the causal effects of the hypothetical tax policy, not the compensated (Hicksian) effects of the policy. See Hendren (2013) for a discussion.

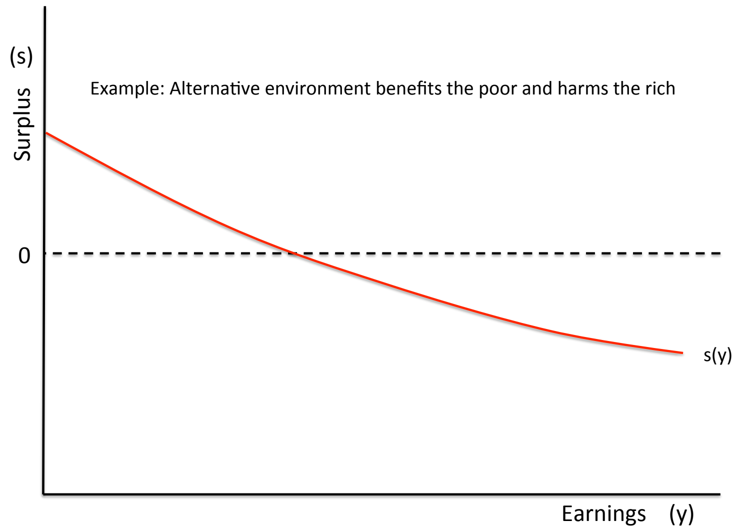

Figure 2: Surplus Example: $s(y)$ represents individuals' willingness to pay (out of their own income) for the alternative environment

status quo and alternative environments; in these cases the two notions inequality deflated surplus will be equivalent to first order, analogous to the first order equivalence of EV and CV in standard consumer theory. In this sense, inequality deflated surplus measures the extent to which everyone can be made better off in the tax-modified alternative environment, relative to the status quo.

The next section develops these ideas more formally, provides the precise first-order statements, and discusses in detail the issues that arise when surplus is heterogeneous conditional on income.

3 Model

This section develops a model of utility maximization subject to nonlinear income taxation in the spirit of Mirrlees (1971) and Saez (2001, 2002). The model is used both to define economic surplus (compensating and equivalent variation) and will also be used to describe the price of transferring resources from one individual to another. This price will then be used to neutralize the interpersonal comparisons involved in the aggregation of surplus. To be consistent with the Kaldor and Hicks environment in which transfers can be individual-specific and heterogeneity may follow arbitrary patterns, I allow each individual to potentially have her own unique utility function and I do not impose restrictions on these functions (such as the Spence-Mirrlees single crossing property).

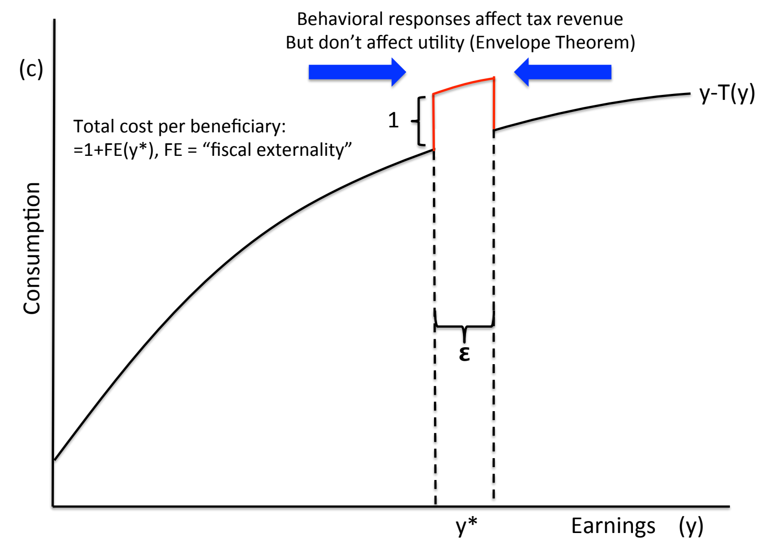

Figure 3: This figure depicts the benefits and costs of making transfers through the income tax schedule. The mechanical beneficiaries of the \$1 tax deduction value the benefits at a dollar. But, the costs of providing this \$1 include both the mechanical costs (\$1) plus the impact of the behavioral response to this policy on the government budget (i.e. the fiscal externality), $FE(y^*)$.

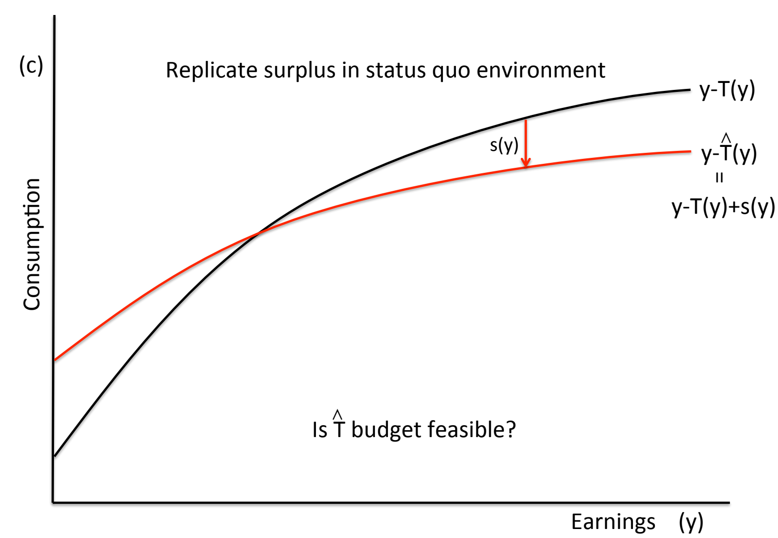

Figure 4: This figure depicts replicating the surplus offered by the alternative environment through modifications to the income tax schedule. To first order, individuals are indifferent between the modified status quo (depicted by the red line) and the alternative environment.

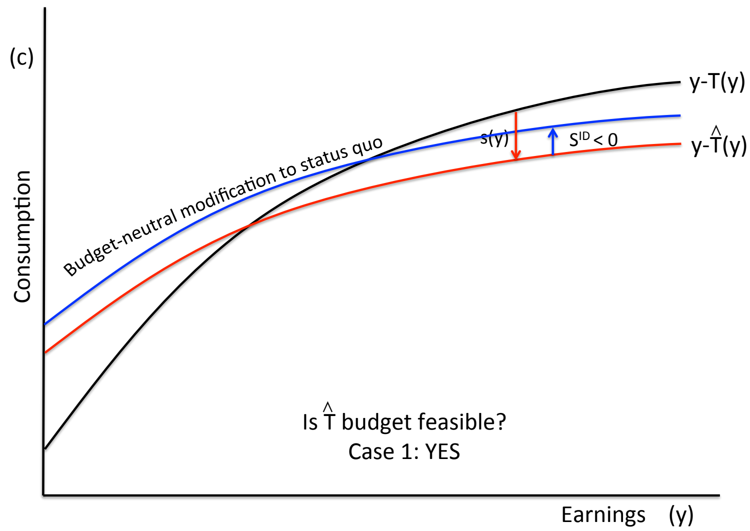

Figure 5: This figure depicts the case when the modification to the status quo is budget feasible, so that closing the budget constraint through a lump-sum tax generates a modified status quo environment that is Pareto superior to the alternative environment.

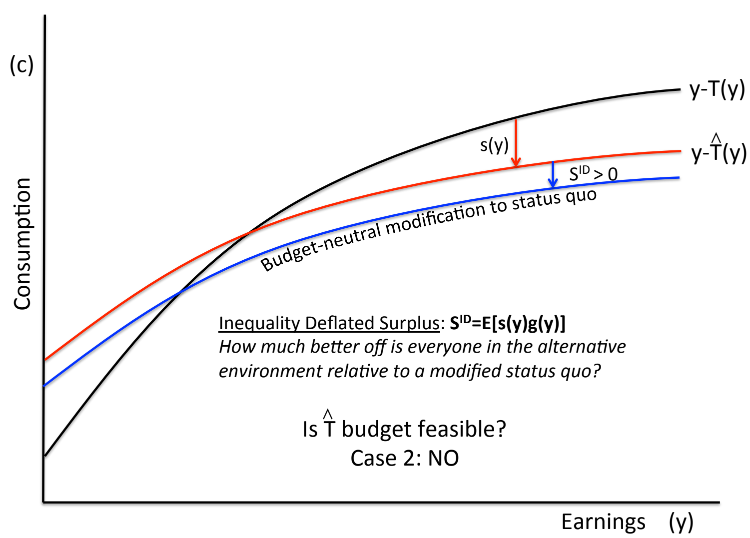

Figure 6: Figure depicts the case when the modification to the status quo is not budget feasible, so that the alternative environment is Pareto superior to the modified status quo world that has the same distributional incidence as the alternative environment.

3.1 Setup

There exists a set of agents indexed by $\theta \in \Theta$, where Θ has measure μ.[21] There is a status quo environment and an alternative environment against which one wishes to compare the status quo. Environments consist of a system of tax policies and utility functions over consumption and earnings.[22] In the status quo environment, type θ chooses consumption, $c(\theta)$, and earnings, $y(\theta)$. I allow each agent to have a potentially different utility function, $u(c, y; \theta)$, over consumption and earnings. Agents maximize utility subject to a budget constraint

$$c(\theta) \leq y(\theta) - T(y(\theta)) + m$$

where $T(y)$ is the taxes paid on earnings y and m is a transfer term that is not contingent upon earnings.[23]

Let $v^0(\theta)$ denote the utility level obtained by type θ in the status quo environment. And, given a utility level v, define the expenditure function $e(v; \theta)$ to be the smallest transfer m that is required for a type θ to obtain utility level v in the status quo environment (i.e. with tax policy $T(\circ)$ and utility functions $u(c(\theta), y(\theta); \theta)$).[24]

In addition to the status quo environment, I consider an alternative environment, a. The notion of "environment" should be interpreted broadly – it can correspond to different policies or laws, different distributions of income, etc. In the alternative environment a, type θ obtains utility $v^a(\theta)$ and has an expenditure function $e^a(v; \theta)$ defined analogously.[25] The goal is to construct a normative criteria under which society prefers one of these environment and to quantify their welfare difference.

To begin, consider the standard equivalent variation measure of the surplus, $s(\theta)$, to type θ from the alternative environment:

$$s(\theta) = e(v^a(\theta); \theta) - e\left(v^0(\theta); \theta\right) \tag{1}$$

This is the amount of additional money a type θ would need in the status quo environment to be just as well off as in the alternative environment. If $s(\theta) > 0$ for all θ, then it must be the case that $v^a(\theta) \geq v^0(\theta)$ for all θ – i.e. that the alternative environment is preferred by all individuals relative to the status quo. In this special case, the Pareto criteria suggests society should prefer the alternative

[21] Formally, I assume (Θ, μ) is a probability space with measure μ (and the usual sigma algebra) so that the entire population is normalized to $\mu(\Theta) = 1$ and one can use the law of iterated expectations by conditioning on subsets of the population, Θ.

[22] Allowing utility functions to vary across environments allow for differences in the disutility of producing earnings, which can be interpreted as productivity differences.

[23] For simplicity, I assume $T(y)$ is the same for everyone. In the empirical implementation, I allow T to vary with individual characteristics, such as the number of dependents, and marital status. See Section 4.1.

[24] Formally, $e = \inf\{m | \sup\{u(c, y; \theta) | c \leq y - T(y) + m\} \geq v\}$. The standard duality result implies that $e\left(v^0(\theta); \theta\right) = m$.

[25] Most of the analysis will not require assuming utility maximization in the alternative environment, but some portions will require describing the marginal cost of taxation in the alternative environment. Here, I assume a structure similar to the status quo whereby individuals maximize a utility function $u^a(c(\theta), y(\theta); \theta)$ subject to a budget constraint, $c(\theta) \leq y(\theta) - T^a(y(\theta)) + m$. The alternative environment is then defined by a different utility function and tax schedule. Note this easily nests models of alternative environments with differences in productivity levels (i.e. disutility of earnings), public goods, taxes, etc.

environment relative to the status quo.

More generally though, alternative environments will not be Pareto ranked relative to the status quo. Some types may have $s(\theta) > 0$ and others $s(\theta) < 0$. As discussed in the introduction, the Kaldor-Hicks compensation principle tests whether aggregate surplus, $\int s(\theta)\, d\mu(\theta)$, is positive.[26] But, this values each types surplus equally. Hence, the Kaldor-Hicks ordering is not distribution-neutral. It opposes any redistribution if it involves a reduction in total surplus.

3.2 The Inequality Deflator

Now suppose that the transfers occur through modifications to the income tax schedule. Following Figure 3, imagine transferring a small tax deduction to those with taxable earnings near y^*. To be precise, let $\eta, \epsilon > 0$ and fix a given income level y^*. Consider providing an additional \$$\eta$ to individuals in an ϵ-region near y^*. Define $\hat{T}(y; y^*, \epsilon, \eta)$ by

$$\hat{T}(y; y^*, \epsilon, \eta) = \begin{cases} T(y) & if\ y \notin \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \\ T(y) - \eta & if\ y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \end{cases}$$

so that $\hat{T}$ provides η additional resources to an ϵ-region of individuals earning between $y^* - \epsilon/2$ and $y^* + \epsilon/2$.

If there were no incentive constraints and the government could target this transfer only to individuals earning between $y^* - \epsilon/2$ and $y^* + \epsilon/2$, then the cost to the government of this transfer would be $\eta\left(F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)\right)$ where $F(y) = \mu(\{\theta | y(\theta) \leq y\})$ is the cumulative distribution of income in the status quo world. However, as observed by Mirrlees (1971), other individuals may choose to alter their earnings towards y^* in order to obtain the transfer of η. By the envelope theorem, these responses do not have first order impacts on utility – the welfare gains are concentrated on the mechanical beneficiaries with incomes in the region $\left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)$.[27] But, these behavioral responses do have first order costs to the government.

To capture this, let $\hat{y}(\theta; y^*, \epsilon, \eta)$ denote the income choice of type θ under the tax schedule

[26] To be formally correct, this is Hicks (1940)'s definition of the principle, which generates a complete and transitive ranking. Kaldor proposed testing whether aggregate compensating variation is positive, where CV uses the expenditure function in the alternative environment:

$$cv(\theta) = e^a(v(\theta); \theta) - e^a(v^a(\theta); \theta)$$

As is well-known, Kaldor's version of the compensation principle is not transitive (Scitovsky (1941)). If environment 1 is preferred to environment 2 and environment 2 is preferred to environment 3, it is not true that environment 1 is necessarily preferred to environment 3. This intransitivity is due to the fact that the expenditure function being used to make comparisons is changing with the alternative environment in question. But, the distinction between EV, CV, and local measures of willingness to pay is second order (Schlee (2013)).

[27] As discussed in Section 3.8, this assumes a local incidence and absence of general equilibrium effects. This provides a natural benchmark, but a more general model would allow tax credits at one part of the income distribution to generate first-order welfare gains at other points of the income distribution. Given empirical evidence on these responses, future work could incorporate these into a more general inequality deflator.

$\hat{T}(y; y^*, \epsilon, \eta)$. Let $\hat{q}(y^*, \epsilon, \eta)$ denote the average cost to the government per mechanical beneficiary:

$$\hat{q}(y^*, \epsilon, \eta) = \frac{-\int \left[\hat{T}(\hat{y}(\theta; y^*, \epsilon, \eta); y^*, \epsilon, \eta) - T(y(\theta))\right] d\mu(\theta)}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)}$$

where the numerator is the cost of the policy and the denominator is the mass of people who receive the mechanical transfer η without any behavioral responses.

Now, consider the marginal cost of providing resources to types near y^*. To calculate this, first I take the derivative of q with respect to η and evaluate at $\eta = 0$. This yields the function $\frac{d\hat{q}(y^*, \epsilon, \eta)}{d\eta}|_{\eta=0}$, which is the marginal cost of providing an additional dollar through the tax code to individuals with earnings in an ϵ-region of y^*. Then, taking the limit as $\epsilon \to 0$, one arrives at the marginal cost to the government of providing an additional dollar of resources to an individual earning y:

$$\lim_{\epsilon \to 0} \frac{d\hat{q}(y, \epsilon, \eta)}{d\eta}|_{\eta=0} = 1 + FE(y) \tag{2}$$

where I assume this limit exists and is continuous in y.[28] In the absence of behavioral responses to the hypothetical tax policies, this marginal cost is \$1 per beneficiary. But, there is an added term in the marginal cost of the transfer which equals the causal impact of the behavioral response to the policy on the government budget, $FE(y)$. This is the "fiscal externality" associated with the behavioral response to the small change to the tax schedule. If the increased transfer causes people to work less and thereby reduces tax revenue, then the fiscal externality is negative; if the policy causes people to work more and thus increases tax revenue, then the fiscal externality will be positive. In general, the size of the fiscal externality is an empirical question and depends on the causal impact of tax changes.

Technical assumption Equation 2 characterizes the marginal cost of providing tax deductions at various points of the income distribution. To use these marginal cost measures at each income level, y, to neutralize distributional comparisons for an entire surplus function across the income distribution, one must assume that a total differentiation property of the government revenue function holds with respect to changes in the tax schedule.[29]

Assumption 1. *Let $y(\theta; T)$ denote the individual's choice of labor earnings in the status quo world when facing tax schedule T. Let $R(T) = \int T(y(\theta; T)) d\mu(\theta)$ denote government revenue. Suppose $\tilde{T}(y) = T(y) + \epsilon \sum_{j=1}^{N} T^j(y)$ for some functions T^j. Let $\tilde{T}^j_\epsilon(y) = T(y) + \epsilon T^j(y)$. Then R is*

[28] Assumption 1 below will imply that this limit exists. In general, this requirement is not very restrictive. But, it would be violated if, for example, there were a mass of people indifferent to earning $y = 0$ and $y = y^*$ so that a small additional transfer induced a massive increase taxes collected at $y = y^*$. Section 3.3 provides a general class of utility functions for which this limit exists.

[29] Because I have allowed for fairly rich heterogeneity, θ, and have not assumed convexity in preferences, changes in the choice of y can be discontinuous in response to small tax changes. I can allow for discontinuous behavioral responses as long as they "average out" when integrating across θ so that the aggregate revenue function is differentiable.

continuously differentiable in ϵ and

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}\right)=\sum_{j=1}^{N}\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\tilde{T}_{\epsilon}^{j}\right)$$

This assumption ensures that the standard tools of calculus characterize government costs when changing the shape of the tax schedule. It is satisfied for most common forms of preferences, such as the broad class of preferences discussed in Subsection 3.3. It would be violated if some types count towards the marginal cost of two different tax movements, as this would lead to a double-counting of marginal costs. This would occur if there were a mass of agents perfectly indifferent between three earnings points in the status quo. Then, providing additional transfers to one of these two points would both induce movement from the other point and thus the sum of the two tax movements would be larger than the combined tax movement. Because Assumption 1 is only a statement about the differentiability of the aggregate government revenue – as opposed to the tax revenue collected from any particular agent – it is weaker than most assumptions imposed in existing literature.

The Inequality Deflator The function $FE(y)$ characterizes how the marginal cost of providing surplus through the tax schedule to those earning near y differs from the mechanical cost of 1. For a given surplus function, $s(\theta)$, I define the inequality deflator as the marginal cost of providing resources to those earning near y normalized by the average marginal cost of providing resources equally across the income distribution.

Definition 1. The **inequality deflator**, $g(y)$, is given by

$$g(y)=\frac{1+FE(y)}{\int(1+FE(y(\theta)))\,d\mu(\theta)}=\frac{1+FE(y)}{E[1+FE(y)]} \tag{3}$$

Inequality deflated surplus is given by

$$S^{ID}=\int s(\theta)\,g(y(\theta))\,d\mu(\theta) \tag{4}$$

The inequality deflator has a straightforward intuition: \$1 of surplus that falls to those earning y can be turned into $g(y)/n$ surplus to everyone in the population (where n is the number of people in the population) through modifications to the income tax schedule. The inequality deflator down-weights (up-weights) surplus if it accrues to individuals to whom it is less (more) costly to redistribute through changes in the income tax schedule.

Multiple Dimensions It is straightforward to verify that the fiscal externality representation of the inequality deflator in equation (3) can be extended to the case when transfers are made based on a multi-dimensional set of characteristics, $\mathbf{X}$, instead of just income, y. In this case, $g(\mathbf{X})=\frac{1+FE(\mathbf{X})}{E[1+FE(\mathbf{X})]}$ could be used to deflate surplus, where $1+FE(\mathbf{X})$ is the marginal cost of providing \$1 of transfers to those with characteristics in an ϵ-region near $\mathbf{X}$. As discussed in Subsection 3.5, this can potentially help provide Pareto comparisons for policies which have heterogeneous surplus conditional on income. But, for most of the paper I focus on the case where transfers are made only conditional on income. This is for two reasons. First, this allows me to draw upon the large body of empirical work studying the behavioral responses to changes in the income tax schedule. Second, Atkinson and Stiglitz (1976) suggests that redistribution through the income tax schedule is, in some cases, sufficient for redistribution.[30]

Negative deflator It may be the case that the inequality deflator is negative, $FE(y) < -1$. This characterizes the existence of Pareto improvements through modifications to the income tax schedule in the status quo environment, and is isomorphic to tests suggested in Werning (2007). Intuitively, $FE(y) < -1$ suggests the existence of a local Laffer effect: the government can increase revenue by providing transfers. In this case, policy recommendations are straightforward and independent of distributional considerations: fix the tax schedule and provide these transfers! But, in the empirical implementation, my results suggest these deflator values are in general non-negative, and hence one cannot find Pareto improvements solely by manipulating the income tax schedule.[31]

3.3 Quantifying the Inequality Deflator

Behavioral responses to these policy changes provide clues about the value of the inequality deflator across the income distribution. At the bottom of the income distribution, existing empirical evidence suggests transfers to the poor through expansions to the EITC schedule induce distortions that increase the cost of the program. For example, Hendren (2013), drawing on studies and summaries in Hotz and Scholz (2003) and Chetty et al. (2013), calculates that a \$1 mechanical increase in EITC benefits has a fiscal cost of around \$1.14. Conversely, Saez et al. (2012) summarize existing literature on the behavioral responses of the top earners to changes in the top marginal tax rate. They suggest the a \$1 mechanical decrease in tax liability through a reduction in the top marginal income tax rate has a fiscal cost of only \$0.50 - \$0.75 because of the induced behavioral responses.

Combining these reduced form estimates, the results imply a shape to the inequality deflator: *surplus in the hands of the rich should be valued less than surplus in the hands of the poor.* Even if one's own social preferences preferred resources in the hands of the rich, a dollar of surplus in the hands of a poor person can be translated to more than a dollar (~\$1.52-\$2.28)) in the hands of a rich person by reducing distortions in the tax schedule.[32] Conversely, a dollar of surplus in the hands of a rich person can only be translated to less than a dollar (~\$0.44-\$0.66) in the hands of a poor person because such movement requires increasing the distortions in the tax schedule.[33] Hence, this leads to

[30] Given individuals, θ, with a full set of observable choices (including income), $\mathbf{X}(\theta)$, and income choice $y(\theta)$, a general statement of the Atkinson and Stiglitz (1976) result is $1 + FE(\mathbf{X}(\theta)) = 1 + FE(y(\theta))$ for all θ. This obviously need not hold in general, but there are well-known weak separability functions on the utility function under which this may hold.

[31] As one incorporates more policy dimensions, $\mathbf{X}$, it may be the case that $FE(\mathbf{X}) < -1$ for some values of $\mathbf{X}$. This characterizes when there exists a modification to the multiple dimensional transfer system that can provide a Pareto improvement.

[32] Note that $\frac{1.14}{0.5} = 2.28$ and $\frac{1.14}{0.75} = 1.52$

[33] Note that $\frac{0.5}{1.14} = 0.44$ and $\frac{0.75}{1.14} = 0.66$

a preference for surplus in the hands of the poor more so than the rich in a ratio of roughly 2-1.

Elasticity Representation While the causal response to changes in the top tax rate and the EITC provide guidance on the size of $FE(y)$ at broad regions of the income distribution, one ideally prefers a more precise estimate of $FE(y)$ at each income level. To do so, I write $FE(y)$ as a function of the shape of the income distribution, the tax schedule, and behavioral elasticities, following Saez (2001), and the more recent inverse optimum literature of Bourguignon and Spadaro (2012), Blundell et al. (2009), Bargain et al. (2011), and Zoutman et al. (2013a,b).[34] Relative to this literature, I make a relatively weak set of assumptions on the unobserved heterogeneity in the model (e.g. I do not assume a uni-dimensional or a small set of discrete types, nor do I assume that a Spence-Mirrlees single crossing condition holds).

Some additional assumptions are required for an elasticity representation of $FE(y)$. In particular, I assume individuals may respond to taxation by choosing to enter the labor force or adjust their labor hours. However, I make the simplification that intensive margin adjustments are continuous in the tax rate. More formally, let $c(y;w,\theta)$ trace out a type θ's indifference curve (in consumption-earnings space) at utility level w, defined implicitly by the standard indifference equation:

$$u(c(y;w,\theta),y;\theta)=w$$

I make the following assumptions.

Assumption 2. *For any $\kappa>0$, let $B(\kappa)=[u(y(\theta)-T(y(\theta)),y(\theta);\theta)-\kappa,u(y(\theta)-T(y(\theta)),y(\theta);\theta)+\kappa]$ denote an interval of width κ near the status quo utility level. Each type θ's indifference curve, $c(y;w,\theta)$, satisfies the following conditions:*

1. (Continuously differentiable in utility) For each $y\geq 0$, there exists $\kappa>0$ such that $c(y;w,\theta)$ is continuously differentiable in w for all $w\in B(\kappa)$
2. (Convex in y for positive earnings, but arbitrary participation decision) For each $y>0$, there exists $\kappa>0$ such that $c(y;w,\theta)$ is twice continuously differentiable in y for all $w\in B(\kappa)$ and $c_y>0$ and $c_{yy}>0$.
3. (Continuous distribution of earnings) $y(\theta)$ is continuously distributed on the positive region $y>0$ (but may have a mass point at $y=0$).

Assumption 2 imposes fairly weak assumptions on the utility function. First, it imposes the standard assumption that indifference curves move smoothly with utility changes. Second, it requires that indifference curves are convex on the region $y>0$. Importantly, this allows for non-convexities on the participation margin, $y=0$ versus $y>0$. So, small changes in the tax schedule can cause jumps

[34] See also Immervoll et al. (2007) who identify the implicit welfare weights that make policymakers indifferent to a proposed tax/transfer policy change, building on Browning and Johnson (1984). Following Kleven and Kreiner (2006); Immervoll et al. (2007), I also incorporate a participation margin decision.

between $y=0$ and $y>0$ (i.e. a participation response). But, the convexity over $y>0$ ensures small changes in the tax schedule only leads to small intensive margin changes in labor supply.[35] Finally, the third part of Assumption 2 is made for simplicity so that I do not require separate formulas for point mass regions of the income distribution.[36] This allows me to characterize the inequality deflator for regions of the tax schedule that are continuously differentiable, which corresponds to the vast majority of points along the income distribution.[37]

Using Assumption 2, one can write the fiscal externality at each point along the income distribution as a function of labor supply elasticities, tax rates, and the shape of the income distribution. Let $\tau(y) = T'(y)$ denote the marginal tax rate faced by an individual earning y. For individuals with $y>0$, the concavity of the utility function implies that the marginal rate of substitution between income and consumption is equated to the relative price of consumption, $1-\tau(y)$:

$$-\frac{u_c}{u_y} = (1-\tau(y(\theta)))$$

I define the average intensive margin compensated elasticity of earnings with respect to the marginal keep rate for those earning $y(\theta)=y$ in the status quo,

$$\epsilon^c(y) = E\left[\frac{1-\tau(y(\theta))}{y(\theta)} \frac{dy}{d(1-\tau)}|_{u=u(c,y;\theta)} | y(\theta) = y\right]$$

which is the percent change in earnings resulting from a percent change in the price of consumption.

I also define the income elasticity of earnings by $\zeta(y)$

$$\zeta(y) = E\left[\frac{dy(\theta)}{dm} \frac{y(\theta) - T(y(\theta))}{y(\theta)} | y(\theta) = y\right]$$

which is the percentage response in earnings to a percent increase consumption.

Finally, let $f(y)$ denote the density of earnings at y. I define the extensive margin (participation) elasticity with respect to net of tax earnings, $\epsilon^P(y)$.

$$\epsilon^P(y) = \frac{d[f(y)]}{d[y-T(y)]} \frac{y-T(y)}{f(y)}$$

With these definitions, Proposition 1 follows.

Proposition 1. *For any point y^* such that $\tau(y^*)$ is constant, the fiscal externality of providing*

[35] This simplifies the representation of the cost of raising tax revenue, since the intensive margin responses will be summarized by local intensive margin elasticities. See Kleven and Kreiner (2006) for a particular utility specification that satisfies Assumption 2 and captures these features of intensive and extensive margin labor supply responses.

[36] If the tax schedule had kinks that generated significant bunching, then one would need to modify the formulas below accordingly. With bunching, equation (2) would continue to characterize the inequality deflator at bunch points, but one would need to derive a different elasticity representation.

[37] Note that the fiscal externality term, $FE(y)$, continues to characterize the inequality deflator at regions of the income distribution where $y(\theta)$ is not continuously distributed.

additional resources to individuals near y^* is given by

$$FE(y^*) = \underbrace{-\epsilon_c^P\left(y^*\right)\frac{T(y^*)-T(0)}{y^*-T(y^*)}}_{Participation\ Effect}\underbrace{-\zeta(y^*)\frac{\tau(y^*)}{1-\frac{T(y^*)}{y^*}}}_{Income\ Effect}+\underbrace{\frac{\tau(y)}{1-\tau(y)}\left(\frac{d}{dy}|_{y=y^*}\left[\epsilon^c(y)\frac{yf(y)}{f(y^*)}\right]\right)}_{Substitution\ Effect} \quad (5)$$

Moreover, for all points y such that $\frac{d}{dy}[\epsilon^c(y)] = 0$, the fiscal externality is given by

$$FE(y^*) = -\epsilon^P(y^*)\frac{T(y^*)-T(0)}{y^*-T(y^*)} - \zeta(y^*)\frac{\tau(y^*)}{1-\frac{T(y^*)}{y^*}} - \epsilon^c(y^*)\frac{\tau(y^*)}{1-\tau(y^*)}\alpha(y^*) \quad (6)$$

where $\alpha(y) = -\left(1+\frac{yf'(y)}{f(y)}\right)$ is the local elasticity of the income distribution (which equals the local Pareto parameter of the income distribution).

Proof. Proof provided in Appendix A □

Proposition 1 is a generalization of the canonical optimal tax formula to the case of multi-dimensional heterogeneity.[38] Consistent with the intuition provided by Saez (2001), Proposition 1 shows that the relevant empirical elasticities in the case of potentially multi-dimensional heterogeneity are the population average elasticities conditional on income.

The fiscal externality associated with providing an additional dollar resources to an individual earning y^* is the sum of three effects. First, to the extent to which the additional transfer induces people into the labor force, this increases tax revenue proportional to the difference between the average taxes received at y^* and the taxes/transfers received from those out of the labor force, $T(0)$. Second, the increased transfer may change the labor supply of those earning y^* due to an income effect. These marginal changes affect the government budget proportional to the marginal tax rate, $\tau(y^*)$. Finally, the transfer might attract other people earning close to y^* who will change their earnings to y^* in order to get the transfer (the substitution effect).

As illustrated in Equation (6), the substitution effect can be thought of as a product of three terms. The elasticity, $\epsilon^c(y)$, measures how responsive people are to the change in the tax; and, the tax ratio, $\frac{\tau(y)}{1-\tau(y)}$, captures the fiscal impact of these responses. However, these substitution costs are the sum of two effects: some people will decrease their earnings towards y^*; others will increase their earnings to y^*. The former effect increases tax revenue; the latter effect decreases tax revenue. Hence, the difference in the earnings density above versus below y^* matters for calculating the fiscal externality. The extent to which the losses outweigh the gains is governed by the elasticity of the income distribution, $\frac{yf'(y)}{f(y^*)}$. When $\frac{yf'}{f}$ is negative, this reduces the cost of providing transfers because more people increase rather than decrease their taxable earnings in order to obtain the transfer. When $\frac{yf'}{f} < -1$, as is the case with the Pareto upper tails in the US income distribution, this lowers the

[38] Zoutman et al. (2013b) provide a derivation of equation 5 for the case when θ is uni-dimensional and distributed over an interval and for which the Spence-Mirrlees single crossing condition holds. The optimal top marginal income tax rate (e.g. Diamond and Saez (2011) and Piketty and Saez (2012)) can be derived by taking the limit as y increases and setting it equal to the mechanical revenue impact of 1.

marginal cost of providing transfers to the point where the distortions induced from tax credits increase total earnings; hence, the marginal cost of providing transfers in this region of the income distribution can be less than 1. In short, even if elasticities and tax rates are constant, the shape of the income distribution plays a key role in determining the marginal cost of taxation.

Summary In sum, the inequality deflator equals the marginal cost of providing transfers through the income distribution if financed equally throughout the distribution, $g(y) = \frac{1+FE(y)}{E[1+FE(y)]}$. The fiscal externality representation of this marginal cost is quite general. Indeed, it has a natural extension to cases where transfers are made conditional on a multi-dimensional choice vector, $\mathbf{X}$, so that $g(\mathbf{X}) = \frac{1+FE(\mathbf{X})}{E[1+FE(\mathbf{X})]}$. Moreover, under the relatively weak additional assumptions that behavioral responses can be represented by elasticities, one can write $FE(y)$ using average taxable income elasticities and the shape of the income distribution and marginal tax rates, as in equations 5 and 6. Given the piece-wise linear nature of the U.S. tax schedule, equation 6 will be the primary method of implementation in Section 4.

3.4 Technical Properties of the Inequality Deflator

The inequality deflator can be used to compare two alternative environments. This subsection provides two precise statements – analogous to compensating and equivalent variation – for what it means to deflate surplus, $s(\theta)$, using the inequality deflator, $g(y)$. Proposition 2 shows that, to first order, deflated surplus is positive if and only if there does not exist a modification to the tax schedule that can make each point of the income distribution better off relative to the alternative environment. Proposition 3 shows that, under an additional assumption, inequality deflated surplus is positive if and only if there exists a modification to the tax schedule in the alternative environment such that each point of the income distribution is, to first order, made better off relative to the status quo. If surplus, $s(\theta)$, does not vary conditional on income, then testing whether inequality deflated surplus is positive amounts to searching for potential Pareto improvements. If surplus is heterogeneous conditional on income, one needs to make additional modifications to the surplus function, discussed in Section 3.5.

To make these "to first order" statements more formally, for any $\epsilon > 0$ define the scaled surplus by $s_\epsilon(\theta) = \epsilon s(\theta)$ and $S_\epsilon^{ID} = \int s_\epsilon(\theta) g(y(\theta)) d\mu(\theta) = \epsilon S^{ID}$. Clearly, $S^{ID} > 0$ if and only if $S_\epsilon^{ID} > 0$. The following proposition compares the alternative environment to a modified status quo environment in which the tax schedule replicates the distributional incidence of the alternative environment.

Proposition 2. *If $S^{ID} < 0$, there exists an $\tilde{\epsilon} > 0$ such that for any $\epsilon < \tilde{\epsilon}$ there exists an augmentation to the tax schedule in the status quo environment that generates surplus, $s_\epsilon^t(\theta)$, that is higher at all points of the income distribution: $E\left[s_\epsilon^t(\theta) | y(\theta) = y\right] > E\left[s_\epsilon(\theta) | y(\theta) = y\right]$ for all y. Conversely, if $S^{ID} > 0$, no such $\tilde{\epsilon}$ exists.*

Proof. See Appendix A.2 for details. The proof follows by constructing a modified tax schedule that provides surplus proportional to $E[s(\theta) | y(\theta) = y]$ at each point along the income distribution and noting that the cost of doing so is non-negative if $S^{ID} < 0$. Conversely, if $S^{ID} > 0$, the such a policy has a negative budget impact. $\square$

When inequality-deflated surplus is positive, to first order one cannot modify the income tax schedule in the status quo world to replicate the average surplus offered by the alternative environment at each point along the income distribution. Conversely, when inequality-deflated surplus is negative, to first order one can replicate the average surplus offered by the alternative environment at each point in the income distribution.[39]

Compensating variation Proposition 2 provides an equivalent variation justification for the use of the inequality deflator. One can also justify the use of the inequality deflator using a compensating variation argument as well. To do so, one needs to make some additional assumptions to describe the redistribution process that occurs in the alternative environment for sufficiently small ϵ.

Suppose for each $\epsilon \in (0, \tilde{\epsilon})$ there exists an alternative environment that yields surplus $s_\epsilon = \epsilon s(\theta)$. Suppose moreover that these ϵ-alternative environments have a structure in which individuals maximize earnings subject to a tax schedule and let $y^\epsilon(\theta)$ denote their income choice.[40] Let $g^\epsilon(y)$ be the inequality deflator in the ϵ-alternative environment ϵ. In order to provide a compensating variation logic for the inequality deflator, I assume that it provides an adequate measure of the cost of redistribution in the ϵ-alternative environments.

Assumption 3. *For sufficiently small $\tilde{\epsilon}$, the inequality deflator in the alternative environment is the same as in the status quo. Specifically, there exists $\tilde{\epsilon}$ such that if $\epsilon \in (0, \tilde{\epsilon})$, then (1) $y^\epsilon(\theta)$ is the same for all types θ that had the same income in the status quo world, $y^\epsilon(\theta) = y^\epsilon(\theta')$ iff $y(\theta) = y(\theta')$, and (2) $g(y(\theta)) = g(y^\epsilon(\theta))$ for all θ. Moreover, Assumption 1 holds for each $\epsilon \in (0, \tilde{\epsilon})$.*

Assumption 3 guarantees that the inequality deflator can be used to measure the cost of redistribution in the alternative environments. If so, the inequality deflator prefers the alternative environment if and only the surplus can be redistributed so that every point of the income distribution receives greater average surplus relative to the status quo.

Proposition 3. *Suppose Assumption 3 holds. If $S^{ID} > 0$, there exists $\tilde{\epsilon} > 0$ such that for any $\epsilon < \tilde{\epsilon}$, there exists an augmentation to the tax schedule in the alternative environment that delivers surplus $s^t_\epsilon(\theta)$ that is on average positive at all points along the income distribution: $E\left[s^t_\epsilon(\theta) | y(\theta) = y\right] > 0$ for all y. Conversely, if $S^{ID} < 0$, then no such $\tilde{\epsilon}$ exists.*

Proof. See Appendix A.3. □

When Assumption 3 holds, positive inequality deflated surplus means one could hypothetically modify the tax schedule in the alternative environment so that the winners compensate the losers. In this sense, inequality deflated surplus entails a logic in the spirit of the Kaldor-Hicks compensation principle. However, instead of searching for potential Pareto improvements, one compares the average surplus at each point of the income distribution. When surplus is heterogeneous conditional on income, the search for potential Pareto improvements is more difficult.

[39] Of course, one should be cautious when applying the deflator to very large changes in environments, as the marginal cost of the required distortionary taxation may begin to differ from the local marginal cost around the status quo environment, embodied in $g(y)$. Moving beyond this first-order approach is an interesting direction for future work.

[40] See Footnote 25 for the statement of this utility maximization structure.

3.5 Heterogeneous Surplus Conditional on Income

If two people earning the same income, $y(\theta)$, have different surplus, $s(\theta)$, then undoing the distributional incidence through the tax schedule will necessarily make one of the two people strictly better off.[41] Fortunately, with a slight modification of the surplus function, one can use the inequality deflator to characterize the existence of local Pareto improvements.

Given the surplus function $s(\theta)$ of interest, I define the min and max surplus at each point of the income distribution. First, for any $\hat{y}$ let $\underline{s}(\hat{y}) = \inf\{s(\theta)\,|y(\theta) = \hat{y}\}$ be the smallest surplus obtained by a type θ that earns $\hat{y}$ (note this number may be negative). Second, let $\overline{s}(\hat{y}) = \sup\{s(\theta)\,|y(\theta) = \hat{y}\}$ be the largest surplus obtained by a type θ that earns $\hat{y}$. The search for local Pareto improvements involves deflating not actual surplus, $s(\theta)$, but rather these min and max surplus functions conditional on income. In particular, let

$$\underline{S}^{ID} = \int \underline{s}(y)\, g(y(\theta))\, d\mu(\theta)$$

and

$$\overline{S}^{ID} = \int \overline{s}(y)\, g(y(\theta))\, d\mu(\theta)$$

If $\overline{S}^{ID} < 0$, then there exists a modification to the existing tax schedule such that everyone locally prefers the modified status quo to the alternative environment.

Proposition 4. *Suppose $\overline{S}^{ID} < 0$. Then, there exists an $\tilde{\epsilon} > 0$ such that, for each $\epsilon < \tilde{\epsilon}$ there exists a modification to the income tax schedule that delivers a Pareto improvement relative to $s_\epsilon(\theta)$. Conversely, if $\overline{S}^{ID} > 0$, there exists an $\tilde{\epsilon} > 0$ such that for each $\epsilon < \tilde{\epsilon}$ any budget-neutral modification to the tax schedule results in lower surplus for some θ relative to $s_\epsilon(\theta)$.*

Proof. See Appendix A.4 □

When $\overline{S}^{ID} < 0$, a change in the tax schedule within the status quo locally Pareto dominates the alternative environment. Clearly, $\overline{S}^{ID} \geq S^{ID}$ so that this is a more restrictive test of whether the status quo should be preferred to the alternative environment.

Conversely, using Assumption 3, one can test whether the alternative environment, modified with a change to the tax schedule, provides a local Pareto improvement relative to the status quo.

Proposition 5. *Suppose Assumption 3 holds. Suppose $\underline{S}^{ID} > 0$. Then, there exists an $\tilde{\epsilon} > 0$ such that, for each $\epsilon < \tilde{\epsilon}$ there exists a modification to the income tax schedule in the alternative environment such that the modified alternative environment delivers positive surplus to all types relative to the status quo, $s^t_\epsilon(\theta) > 0$ for all θ.*

Proof. See Appendix A.5 □

[41] For another example, suppose an alternative environment offers a surplus of \$20 to one person earning \$40K and a surplus of -\$10 to another person also earning \$40K. Then the inequality deflator would ask whether one can modify the tax schedule to provide \$5 of surplus to those earning \$40K. Of course, this \$5 would not sufficiently compensate the individual with -\$10 in surplus and hence the inequality deflated surplus would not correspond to a potential Pareto improvement.

In general, it can be the case that $\overline{S}^{ID} > 0 > \underline{S}^{ID}$, so that the potential Pareto criterion cannot lead to a sharp comparison between the status quo and the alternative environment. But in many applications, such as the comparisons of income distributions in Section (5.1) and Section (5.2), the surplus will not vary with θ conditional on income $y(\theta)$. Therefore, $\underline{s}(y) = \overline{s}(y) = s(\theta)$ whenever $y(\theta) = y$. In these cases, applying the inequality deflator is equivalent to searching for potential Pareto improvements through modifications to the income tax schedule, in the spirit of Kaldor (1939) and Hicks (1939, 1940).

Corollary 1. *Suppose $s(\theta)$ does not vary with θ conditional on income, $y(\theta)$ (i.e. $s(\theta) = \tilde{s}(y(\theta))$). Then, $\overline{S}^{ID} = \underline{S}^{ID} = S^{ID}$.*

3.6 Relationship to Social Welfare Function

If a social welfare function exists, the inequality deflator equals the average implicit social marginal utilities of income that rationalize the status quo tax schedule as optimal. To see this, let $\chi(\theta)$ denote the social marginal utility of income of individual θ, normalized so that $E[\chi(\theta)] = 1$. In the social welfare function approach, ratios of social marginal utilities of income, $\frac{\chi(\theta_1)}{\chi(\theta_2)}$, characterize the social willingness to pay to transfer resources from θ_2 to θ_1 and provide a generic local representation of social preferences (Saez and Stantcheva (2013)).

Proposition 6. *Suppose the income tax schedule in the status quo, $T(y)$, is maximizes social welfare and let $\chi(\theta)$ denote the local social marginal utilities of income. Then, the inequality deflator, $g(y)$, equals the average social marginal utilities of income for those earning $y(\theta) = y$,*

$$g(y) = E[\chi(\theta) | y(\theta) = y]$$

Proof. Given a tax function $\hat{T}(y; y^*, \epsilon, \eta)$, let $\hat{v}(\theta, \epsilon, \eta)$ denote the utility to type θ. By the envelope theorem, we have

$$\frac{d\hat{v}}{d\eta}|_{\eta=0} = \begin{cases} 0 & if\ y \notin \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \\ \frac{\partial v(\theta)}{\partial m} & if\ y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \end{cases}$$

so that the impact on the social welfare function is $\int \chi(\theta) 1\left\{y(\theta) \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)\right\} d\mu(\theta)$, where $\chi(\theta)$ equals $\frac{\partial v(\theta)}{\partial m}$ multiplied by the local social welfare weight. Taking the limit as $\epsilon \to 0$, we have that the benefit of a small increase in η is $E[\chi(\theta) | y(\theta) = y]$; moreover, by definition the cost of a small increase in η is $g(y)$. Optimality of the tax code implies that the welfare benefit per unit cost is equated for all y:

$$\frac{E[\chi(\theta) | y(\theta) = y_1]}{E[\chi(\theta) | y(\theta) = y_2]} = \frac{g(y_1)}{g(y_2)}$$

Finally, note that $g(y) = \frac{E[\chi(\theta)|y(\theta)=y]}{E[\chi(\theta)|y(\theta)=y_2]} g(y_2)$, so that $E[g(y)] = \frac{E[\chi(\theta)]}{E[\chi(\theta)|y(\theta)=y_2]} g(y_2)$. Now, by construction $E[g(y)] = 1$ and $E[\chi(\theta)] = 1$, so replacing notation of y_2 with y yields $g(y) = E[\chi(\theta) | y(\theta) = y]$. □

If the tax schedule maximizes a social welfare function, then the inequality deflator weights surplus by the average social marginal utilities of income at each income level. In this sense, it provides lower weight to those whose income levels have lower social marginal utilities and higher weight to those with higher social marginal utilities of income.

In this sense, the inequality deflator is related to growing literature studying the inverse optimum program (Saez (2002); Bourguignon and Spadaro (2012); Blundell et al. (2009); Bargain et al. (2011); Zoutman et al. (2013a)). This literature seeks to characterize the marginal cost of taxation along the income distribution (i.e. $g(y)$ in the present notation) and used it to infer the social marginal utilities of income of those who set the tax schedule. While a primary motivation for this literature has been to better understand the implicit preferences arising from the political process, the present analysis provides formal justification for using the inequality deflator as a general method to weight surplus across the income distribution. Here, the Kaldor-Hicks logic provides such justification *even if one's own social preferences differ from those that rationalize the status quo tax schedule as optimal.* If the alternative environment delivers greater inequality-deflated surplus, then it provides greater surplus to every point of the income distribution than can be obtained through modifications through the income tax schedule.

Heterogeneity Conditional on Income The mapping from social marginal utilities of income to the nonlinear tax schedule cannot be inverted when the social marginal utilities of income can vary conditional on income. Intuitively, two different people with the same income may have different social preferences. In such instances, inequality deflated surplus does not correspond to a measure of the change in implicit social welfare, even if one wanted to use the implicit social welfare function that rationalized the tax schedule as optimal. The difference between inequality deflated surplus, $E[s(\theta)g(y(\theta))]$, and implicit social welfare, $E[s(\theta)\chi(\theta)]$, equals the covariance between the social marginal utility of income and surplus, conditional on income.

Corollary 2. *The surplus weighted by the social marginal utilities of income differs from the inequality deflated surplus of the alternative environment by the covariance of surplus and social marginal utilities of income, conditional on income:*

$$\begin{aligned} E[s(\theta)\chi(\theta)] &= E_y\left[E_{\theta|y}[s(\theta)|y(\theta)=y]\,E[\chi(\theta)|y(\theta)=y]\right] + E_y\left[cov_{\theta|y}(\chi(\theta), s(\theta)|y(\theta)=y)\right] \\ &= E[s(\theta)g(y(\theta))] + E_y\left[cov_{\theta|y}(\chi(\theta), s(\theta)|y(\theta)=y)\right] \end{aligned}$$

In particular, if either (a) surplus does not vary conditional on income or (b) social marginal utilities of income do not vary conditional on income, then the inequality deflated surplus corresponds to weighting surplus using the social marginal utilities of income that rationalize the status quo tax schedule as optimal.

Corollary 2 highlights an important caveat in using the inequality deflator for making welfare comparisons. If a policy provides systematic surplus to individuals of a socially-valued type, conditional on income, the inequality deflated surplus will be lower than would be obtained if one used the implicit social marginal utilities of income that rationalize the tax schedule as optimal. For example, it could be the case that Medicaid is a more costly method of redistribution than the income tax. But, Medicaid provides greater surplus to socially valued groups (e.g. the sick) conditional on their income.

More generally, the inequality deflated surplus does not bound the potential social welfare of the alternative environment, as illustrated in the following Corollary.

Corollary 3. *Consider a policy where $s(\theta) = \epsilon$ for the entire population except those whose earnings are in a non-trivial region of the income distribution, $y(\theta) \in [y^*, y^* + a]$. For those with incomes in this region, half have surplus $s(\theta) = -\epsilon$ and the other half has surplus $s(\theta) = \epsilon$, where surplus is independent of income. As long as $g(y)$ is bounded (which will be the case empirically), such a policy delivers a potential Pareto improvement for sufficiently small ϵ and a. But, for any $M > 0$ there exists two sets of positive weights, $\chi_1(\theta)$ and $\chi_2(\theta)$, such that:*

1. *The conditional mean of $\chi_1(\theta)$ and $\chi_2(\theta)$ equals the inequality deflator (and hence can rationalize the tax schedule as optimal)*

$$E\left[\chi_1(\theta) \,|y(\theta) = y\right] = E\left[\chi_2(\theta) \,|y(\theta) = y\right] = g(y)$$

2. *Surplus weighted by $\chi_1(\theta)$ is arbitrarily large:*

$$E\left[\chi_1(\theta) \, s(\theta)\right] > M$$

3. *Surplus weighted by $\chi_2(\theta)$ is arbitrarily small*

$$E\left[\chi_2(\theta) \, s(\theta)\right] < -M$$

Proof. Follows straightforwardly by assuming the conditional distribution of $\chi(\theta)$ given income $y(\theta)$ follows a Pareto distribution with shape parameter sufficiently close to 1, a scale parameter chosen to match the conditional mean, $g(y)$, and a correlation with $s(\theta)$ that is either positive (case 2) or negative (case 3).The intuition is that the inequality deflator pins down the conditional mean, but not the variance, of the implicit welfare weights that rationalize the tax schedule as optimal. □

The Corollary shows that one can have an alternative environment that generates a potential Pareto improvement, but the impact on social welfare of the policy itself cannot be bounded even if that same social welfare function rationalizes the tax schedule as optimal. So, although the inequality deflator has a strong link to the optimal inverse program, it cannot be used to estimate, or even bound, the social welfare impact of alternative environments in the presence of heterogeneity conditional on income. However, it can continue to be used to characterize the existence of potential Pareto improvements using the functions $\overline{S}^{ID}$ and $\underline{S}^{ID}$.

3.7 Dealing with Heterogeneity in Practice

When surplus is heterogeneous conditional on income, it may be the case that $\overline{S}^{ID} > 0 > \underline{S}^{ID}$. In this case, there does not exist a modification to the tax schedule in the alternative or status quo environment that can render a Pareto comparisons between the status quo and alternative environment. Here, there are several options. First, one could bias the status quo, choosing the alternative environment iff $\underline{S}^{ID} > 0$. Of course, this might be overly conservative. Second, one can use inequality deflated average surplus, S^{ID}, and decide if the alternative environment brings sufficient benefits to each point of the income distribution to warrant the lack of Pareto improvement. This approach of course violates the Pareto principle, but may be a useful application of the deflator in cases with important sources of heterogeneity conditional on income.

Third, one could consider additional compensation instruments, such as capital taxation, commodity taxation, Medicaid eligibility, etc. Intuitively, when $\overline{S}^{ID} > 0 > \underline{S}^{ID}$, the income tax alone is too blunt an instrument to conduct compensating transfers. For example, if surplus is a function of both health and income, one could imagine making compensating transfers through modifications to both income and Medicaid / Medicare generosity and eligibility. Here, one requires estimates of $FE(\mathbf{X})$ (e.g. if $\mathbf{X} = (y, m)$ where m is Medicaid expenditures m, one requires the causal effect of the behavioral response to a transfer directed towards those not only with income y but also with Medicaid expenditures m. The key requirement is empirical estimation of the fiscal externalities.

Finally, one can consider policies that have smaller variations in surplus conditional on income. Intuitively, it is likely easier to find Pareto improvements for policies of the form "approve mergers of type X" as opposed to policies of the form "approve merger X", since the willingness to pay can be thought of as ex-ante to the set of mergers that will be approved. The inequality deflator is well-suited to addressing comparisons where the key source of heterogeneity is income, as will be the case in the examples in Section 5.

3.8 Discussion of Limitations

The previous section highlighted the difficulties that arise when surplus is heterogeneous conditional on income. Before turning to the empirical implementation, I want to highlight a couple of additional issues that could be incorporated into the inequality deflator, but are currently excluded.

GE effects and spillovers By writing income, $y(\theta)$, into the utility function and not including the income earnings of anyone else, I implicitly rule out general equilibrium effects associated with changes to tax policies. For example, if increasing taxes on the rich leads to lower wages for the poor, as in Rothschild and Scheuer (2013), then this would increase their marginal disutility of earning a given income level. Conversely, if increasing taxes on the rich leads them to conduct less rent-seeking activities that harm the earnings of the poor, this would decrease their marginal disutility of earning a given income level. My baseline assumption is that these general equilibrium effects and spillovers are insignificant relative to the partial equilibrium behavioral responses. But future work could incorporate these responses.

Tax avoidance It is widely believed that much of the behavioral response to taxation, especially at the upper regions of the income distribution, is due to avoidance behavior as opposed to changes in labor supply (Saez et al. (2012)). In general, the fiscal externality, $FE(y)$, includes responses both from changes to costly avoidance activities or to changes in labor supply. As noted by Feldstein (1999), one need not know why taxable income is changing as long as individuals are optimizing when considering their avoidance choices and avoidance involves real resource costs. However, if individuals are not optimizing – as in Chetty (2009) – or if the costly avoidance activities have externalities (e.g. charitable donations), then these avoidance activities may need to be incorporated. I leave the incorporation of these effects for future work.

4 Estimating the Inequality Deflator

I estimate the inequality deflator across the income distribution by (a) calibrating estimates of behavioral elasticities from existing literature and assessing a robustness to a range of estimates and (b) estimating the joint distribution of income and marginal tax rates using data from the universe of U.S. income tax returns.

4.1 Joint Distribution of Income and Tax Rates

For the shape of the income distribution and tax rates, I use tax return data from the 2012 IRS Databank. The Databank contains de-identified information derived from a virtually complete set of U.S. income tax returns. The use of population tax data allows one to account for a salient feature of the U.S. tax schedule: filers face different tax schedules depending on their characteristics, such as marital status and number of dependents.

To capture these differences, assume there exists a set of tax schedules, $T_j(\circ)$ for $j \in \{1, 2, ..., J\}$ and assume individual of type θ faces tax schedule $j(\theta)$.[42] In this case, the aggregate fiscal externality from providing additional resources to all those earning near y^* across all tax schedules is given by

$$FE(y^*) = E_j[FE_j(y^*)]$$

where $FE_j(y^*)$ is given by equation (5) (or equation (6) if the marginal tax rate is constant) and the expectation is a population average across tax schedules conditional on income level y^*. So, if there are multiple tax rates at a given level of taxable income, one requires estimates of the local Pareto parameter of the income distribution corresponding to each tax rate, α_j. The population data allow for a nonparametric estimate of the shape of the income distribution conditional on the tax schedule.

[42] For simplicity, I assume individuals do not choose their tax schedule. This assumes changes in taxes do not affect the number of kids or choice of marital status. Future work could incorporate an elasticity of the choice of tax schedule in response to targeted transfers at different points along the income distribution.

Income definitions In keeping with the spirit of using the tax code to redistribute income, I define income using IRS income definitions. I define y to be the tax return's ordinary income in 2012, which equals taxable income (f1040, line 43) minus income not subject to the ordinary income tax (long-term capital income (line 13) and qualified dividends (line 9b)). Ordinary income is primarily comprised of labor income, but also includes deductions for things like the number of children and charitable donations.

In addition to the IRS definition of ordinary income, it will be useful to have a measure of income that more closely resembles a measure of household or family income. For this, I follow Chetty et al. (2013) in defining family income as the sum of pre-tax labor and capital income. Formally, it is equal to adjusted gross income plus tax-exempt interest and the non-taxable portion of social security benefits. For married families, I sum this measure across spouses.

Sample I use the universe of 2012 returns from primary filers aged 25-60 and their married spouses, if applicable.[43] Following Chetty et al. (2013), I restrict the sample to households with positive family income. Details of the sample and data construction are provided in Section B and Appendix Table I presents the summary statistics. The resulting sample has roughly 100 million filers.

Tax Rates To each tax return, I assign the marginal tax rate faced by the 2012 federal income tax schedule. The federal rate schedule on ordinary income provides the marginal tax rate for the filer as long as s/he did not have any additional tax credits, such as the earned income tax credit, and was not subject to the alternative minimum tax. If the individual was subject to the alternative minimum tax (AMT), I record their marginal tax rate at the 28% AMT level. If the filer received EITC, I add the marginal tax rate faced on the EITC schedule using information on the number of EIC-eligible children (filed in the tax return), filing status, and the size of the EITC benefit claimed. This provides a precise measure of the federal marginal tax rate faced by each filer on an additional dollar of ordinary income.

In addition to federal taxes, I account for state and local taxes. For state taxes, I assume a constant tax rate of 5% and account for the fact that state taxes are deductible from federal tax liability when calculating the total marginal tax rate. For Medicare and sales taxes, I follow Saez et al. (2012) and assume a 2.9% tax rate for Medicare and a 2.3% sales tax rate. Finally, some states provide additional EITC benefits. To account for this, I assume a 10% "top-up" EITC rate for EITC filers.[44]

[43] I exclude individuals below age 25 because of the likelihood they still live at home and are part of another household. I exclude people above 60, the age at which many begin exiting the labor force and begin collecting unearned income such as social security income or savings withdrawals.

[44] Choosing alternative values for the state tax rates or the EITC rates do not significantly alter the estimates of the deflator. It turns out that the primary driver of the shape of the deflator is the Pareto parameter combined with the assumption of a constant elasticity. As discussed below, the precise characterization of the inequality deflator lines up with the coarse evidence discussed in Section 3.3.

4.2 Elasticity Calibration

For the elasticities, I construct a baseline specification and two additional specifications with smaller and larger elasticities. I begin with the baseline specification.

To start, I assume away intensive margin income effects, consistent with a large literature suggesting such effects are small (Gruber and Saez (2002); Saez et al. (2012)). Second, I draw upon the large literature studying the behavioral response to the EITC. Chetty et al. (2013) calculates elasticities of 0.31 in the phase-in region (income below \$9,560) and 0.14 in the phase-out region (income between \$22,870 and \$43,210). Using the income tax return data, I assign these elasticities to EITC filers in these regions of the income distribution. Second, for filers subject to the top marginal income tax rate, I assign a compensated elasticity of 0.3. This is consistent with the midpoint of estimates estimated from previous literature studying the behavioral response to changes in the top marginal income tax rate (Saez et al. (2012)). Finally, for those not on EITC and not subject to the top marginal income tax rate, I assign a compensated elasticity of 0.3, consistent with Chetty (2012) who shows such an estimate can rationalize the large literature on the response to taxation.

In addition to these intensive margin responses, there is also significant evidence of extensive margin behavioral responses, especially for those subject to the EITC. This literature suggests EITC expansions are roughly 9% more costly to the government due to extensive margin behavioral responses.[45] Therefore, I assume the participation effect in equation (6) is equal to 0.09 for income groups subject to the EITC.

Above the EITC range, there is mixed evidence of participation responses to taxation. Liebman and Saez (2006) find no statistically significant impact of tax changes on women's labor supply of women married to higher-income men. Indeed, higher tax rates can reduce participation from a price effect but increase participation due to an income effect. As a result, I assume a zero participation elasticity for those not subject to the EITC for my baseline specification.

Alternative Specifications In addition to the baseline specification, I also consider two alternative specifications. First, I consider a high elasticity specification that increases the intensive margin compensated elasticity to 0.5; Second, I consider a low elasticity specification that reduces the intensive margin compensated elasticity to 0.1. In both alternative specifications, I continue to assume no income effects and a participation effect of 0.09 for EITC filers.

4.3 Estimation of Shape of Income Distribution

Translating these elasticities into a measure of the fiscal externality in equation 6 requires estimating the elasticity of the density of the income distribution at each income level, conditional on each marginal tax rate, τ_j. In practice, there are many factors that lead two filers with the same taxable income to face different tax rates, such as filing status, EITC eligibility, and subjectivity to the AMT. The combination of filing status, EITC status (which includes marital status plus number of qualifying

[45] See Hotz and Scholz (2003) for a summary of elasticities and Hendren (2013) for the 9% calculation.

dependents), and AMT subjectivity leads to more than 100 different tax schedules, each with numerous different marginal tax rates along the income distribution.

I use the universe of income tax returns from 2012 to non-parametrically estimate the shape of the income distribution, $\alpha_j(y)$, conditional on the marginal income tax rate in a manner that allows this shape to vary arbitrarily across tax schedules for the majority of filers. The details of the estimation are provided in Appendix B. Given these estimates, I construct the measure of the $FE(y)$ using equation 6.

Although the precise calculation of $FE(y)$ relies on estimates of $\alpha_j(y)$ for each tax rate, Figure 7 provides a picture of the mean value of α at each point in the income distribution (aggregating across tax schedules). Consistent with findings in existing literature, the average $\alpha_j(y)$ reaches around 2.5 in the upper regions of the income distribution. This corresponding to an elasticity of the income distribution of 1.5, consistent with findings in previous literature (Diamond and Saez (2011) and Piketty and Saez (2012)). However, the average $\alpha_j(y)$ exhibits considerable heterogeneity across the income distribution. It is negative below the 60th percentile of the income distribution.[46] This means that higher compensated elasticities correspond to higher costs of providing surplus at these points of the income distribution. Conversely, above the 60th percentile, the average $\alpha_j(y)$ is positive, which means higher compensated elasticities correspond to lower costs of providing surplus to upper regions of the income distribution. Putting these findings together, the shape of the income distribution suggests that, even if tax rates and elasticities were constant across the income distribution, the cost of providing \$1 of surplus through the tax schedule will be lower at the top of the income distribution than at the bottom of the income distribution.[47]

4.4 Results

Figure 8 presents the baseline estimates of the inequality deflator. The results suggest the cost of providing an additional dollar to an individual earning y dollars is higher for low-income than for high-income households. Taking a dollar from the top of the distribution can generate around 0.65/1.15=0.57 of welfare to someone at the bottom of the distribution, consistent with the 0.44-0.66 discussed in Section 3.3. However, the results also illuminate the heterogeneity in this cost across the income distribution. The deflator is concave for most of the income distribution. This means that the relative cost of redistribution from median earners to the poor is lower than the relative cost of redistribution from top earners to median earners. Moreover, the deflator exhibits an increase above the 98th percentile. This suggests that surplus in the hands of the very rich is valued higher than surplus in the hands of the 98th percentile of the income distribution. From a positive perspective, this suggests the current U.S. income tax schedule implicitly values income more in the hands of the

[46] This is consistent with the findings of Werning (2007) who estimates the marginal cost of taxation using the SOI public use file.

[47] The fact that $\alpha(y)$ declines in the very top portions of the income distribution is also shown in Saez (2001) and is not an artifact of the estimation method. For high values of income, y, I estimate $\alpha(y)$ by assuming a Pareto tail (Saez (2001)). As discussed in Appendix B and shown in Figure 16, the non-monotonicity occurs at a point that differs is not driven by the switch between estimating $\alpha(y)$ using the local polynomials to the approach that assumes a Pareto upper tail.

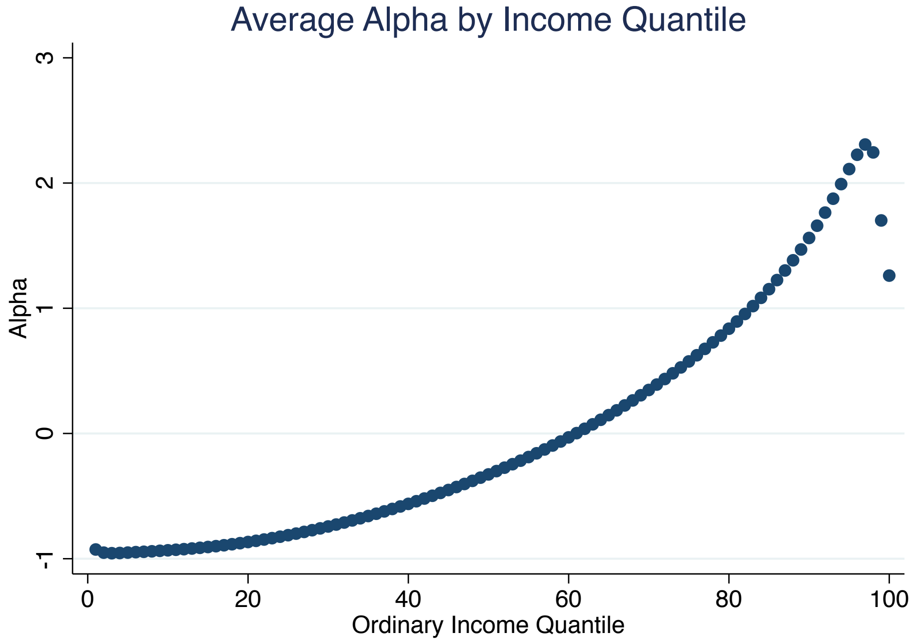

Figure 7: Alpha, $\alpha(y) = -\left(1 + \frac{yf'(y)}{f(y)}\right)$, where $f(y)$ is the density of the income distribution

top 1% than near the 98th percentile.

Figure 9 assesses the robustness of the estimates to alternative elasticity specifications. Valuing surplus in the hands of the rich less than the poor is robust across all specifications. Higher elasticities imply higher costs from behavioral distortions due to taxation. This leads to higher (lower) costs of redistributing from rich to poor. Also, Figure 9 provides insight into the robustness of the increase in $g(y)$ at the top of the income distribution. If one believes that the elasticity of taxable income increases from 0.3 to 0.5 around the 98th-99th percentile, then one cannot rule out that the deflator is monotonically decreasing in income.

Finally, across all specifications, the estimates of the deflator are non-negative. This suggests the existing tax schedule is Pareto efficient: there does not exist an augmentation to the tax schedule that can make everyone better off.[48]

4.5 Family/Household Income Re-weighting

This IRS ordinary income definition corresponds precisely to the conceptual experiment of increasing the marginal tax rate on labor earnings in order to redistribute income. But, it suffers two drawbacks as a unit of measurement for welfare comparisons. First, tax returns can correspond to different numbers

[48]This contrasts with the findings of Bourguignon and Spadaro (2012) using survey data in France that suggests the implicit weights are non-parisian (i.e. $g(y) < 0$) in the upper tail of the income distribution. Although it is of course possible that the French taxation system is above the Laffer curve, one concern with estimates from survey data is that they have trouble estimating the Pareto parameter in the upper tail of the income distribution, leading to an over-estimate of the fiscal externality associated with taxation.

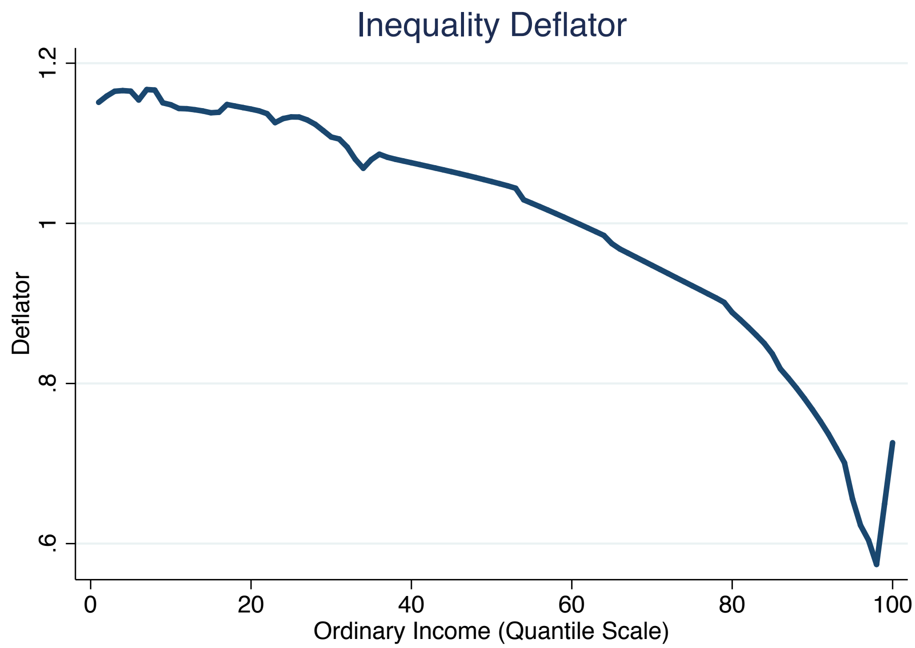

Figure 8: Inequality Deflator, $g(y)$. \$1 of surplus falling to those earning y can be turned into $g(y)/n$ surplus to everyone.

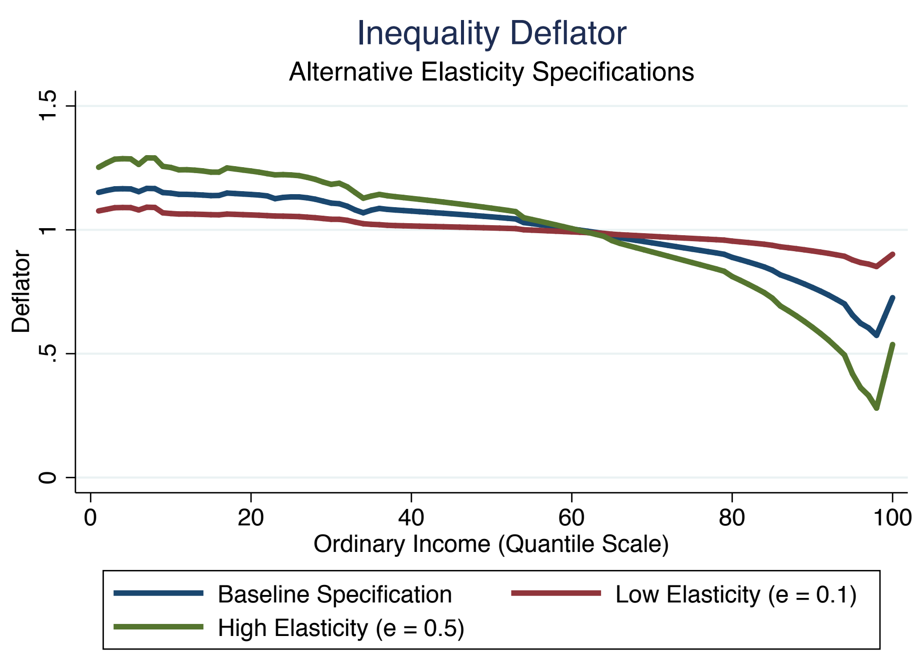

Figure 9: Inequality Deflator, $g(y)$, under alternative elasticity scenarios. Low elasticity specification uses constant compensated elasticity of 0.1; high elasticity specification uses constant compensated elasticity of 0.5.

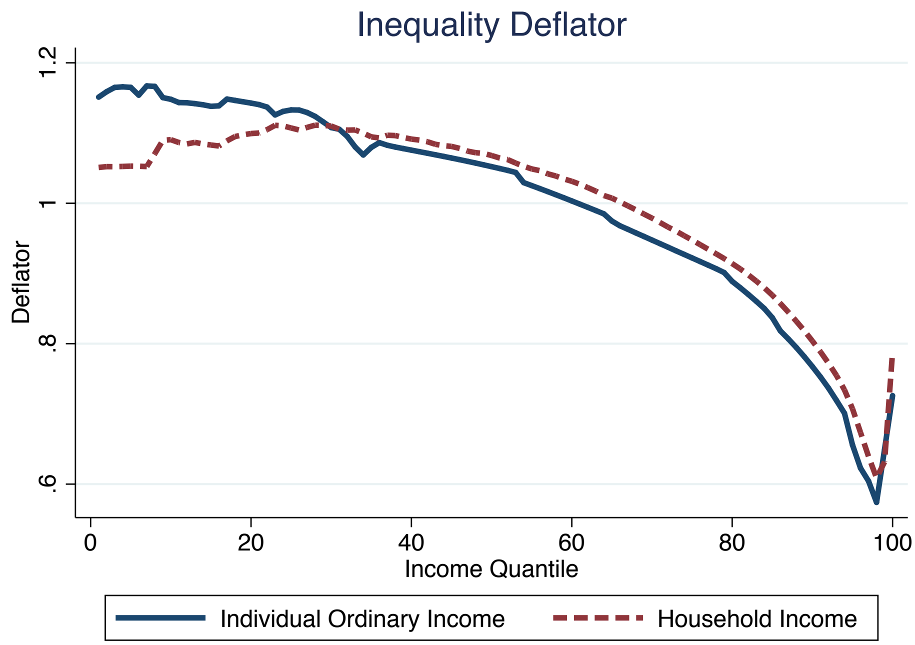

Figure 10: Inequality deflator for individual ordinary income and household income. The blue line is equivalent to Figure 8; the dashed red line presents deflator corresponding to the distribution of household income using the re-weighting in equation 8.

of people: some people file jointly as married, others file as as singles. Moreover, the ordinary income definition adopted by the IRS does not directly correspond to notions of household income often collected from other sources such as the U.S. Census Bureau.

For a simplified application of the deflator to distributions of household or family income, I construct a modified deflator that corresponds to family income comparisons. For a given type θ, let $h(\theta)$ denote their household or family income as defined in Section (4.1). For each level of household income, h, I compute the average value of the inequality deflator for people with household income equal to h:

$$g^H(h) = E\left[g\left(y\left(\theta\right)\right)|h\left(\theta\right) = h\right] \tag{7}$$

so that g^H is the average value of the inequality deflator for individuals with h dollars of family income.

To estimate g^H in equation (7), I construct 100 quantile bins of both household and individual ordinary income and construct observation counts in each household/individual income combination. I then construct the conditional distribution of ordinary income given household income, $w_{y,h}$, so that $g^H(h)$ is given by the weighted sum:

$$g^H(h) \approx \sum g\left(y\left(\theta\right)\right) w_{y,h} \tag{8}$$

Intuitively, $g^H(y)$ is the average cost of providing transfers to those with household income y using modifications to the ordinary income tax schedule.

Figure 10 presents both the ordinary income and household income-based measures of the inequality deflator. As one would expect, the deflator is flatter when viewed as a function of household income quantiles. This is because individuals at the extreme of the income distribution are more likely to be paired with those towards the interior, generating a slightly flatter slope. However, the general pattern remains quite similar. There are higher costs of redistributing from rich to poor and lower costs of redistributing from poor to rich.

5 Applications: Income Distributions

[Using transfers], "it is always possible for the Government to ensure that the previous income-distribution should be maintained intact" (Kaldor (1939)).

I apply the inequality deflator to neutralize interpersonal comparisons when comparing income distributions over time and across countries. Motivated in part by the expenditure function analysis of Kaldor and Hicks original work, it is well-recognized that one should adjust for differences in prices when making normative comparisons of income (e.g. inflation adjustments over time and purchasing-power parity adjustments across countries). Here, I use the inequality deflator to adjust for differences in the distribution of income. I begin with an analysis of changes in the U.S. income distribution over time; I then explore cross-country differences in income distributions.[49]

To compare income distributions, one needs to define a conceptual experiment that yields the surplus function. One could imagine a wide set of possible experiments where people who are at the top of the distribution stay at the top of the distribution in the alternative environment; conversely one could imagine an experiment where people at the top switch with those at the bottom. More generally, to each individual at quantile α in the status quo world, one can be assigned to a quantile $r(\alpha)$ in the alternative environment, where $r(\alpha)$ is a permutation function on $[0, 1]$.[50] This generates surplus,

$$s^r(\alpha) = Q_a(r(\alpha)) - Q_0(\alpha)$$

While there are many potential ways to define surplus for the comparison of two distributions, I will define the surplus experiment as one that maintains quantile stability, $r(\alpha) = \alpha$, so that each person's relative position in the income distribution is maintained intact.[51] This definition minimizes the extent to which individuals are shuffled around in the economy and has a several nice properties. First, it generally reduces the size of each individual surplus, which helps make the first-order approximation for the marginal cost of taxation more appropriate. Second, if one wishes to view the inequality deflator as a social welfare function in the spirit of the optimal inverse literature, then choosing $r(\alpha) = \alpha$ limits

[49] There is a large literature focused on methods for measuring differences in income distributions (see Cowell (2000) for a review). These methods often assess the degree of inequality using theoretically-motivated orderings such as the Lorenz ordering and second order stochastic dominance.

[50] Formally, $r(\alpha)$ is a bijection on $[0, 1]$ and if U is a uniform random variable on $[0, 1]$, then $X = r(U)$ is also a uniform random variable.

[51] In principle, one should use the after-tax income distribution so that differences in the distributions correspond to differences in the net incomes that would be obtained by that quantile in the two environments.

the extent to which social welfare weights would change in the conceptual experiment.[52] Third, and related to the first two points, choosing $r(\alpha) = \alpha$ minimizes the estimated surplus of the status quo relative to the alternative environment (shown in Appendix A.6). Intuitively, having $r(\alpha) \neq \alpha$ adds an additional redistributive component to the distributional comparison that has value because of the desire for redistribution; but is arguably not relevant for making distributional comparisons.

By deflating differences in the quantiles of the income distribution, I abstract from any differences in the value of leisure or public goods across the income distribution. This provides a useful benchmark, but is an important assumption. For example, if one believes that the European countries in the examples below have a more (less) equal allocation of leisure or public goods than the U.S., then the comparisons below will under-state (over-state) the adjustments needed for distributional comparisons. Future work could take estimates of the value of leisure and public goods across the income distribution and directly incorporate them into the surplus function.

5.1 Income Growth in the U.S.

It is well-known that income inequality in the U.S. has increased in recent decades, especially at the top of the distribution (Piketty and Saez (2003)). Figure 11 plots several quantiles of the household after-tax income distribution over time using data from the Congressional Budget Office (CBO) from 1979-2009.[53] As is well-known, incomes have increased significantly in the top portions of the income distribution, especially the top 20% and top 1%; in contrast, income for the bottom 80% has experienced smaller growth.[54]

Here, I use the inequality deflator to calculate how much richer all points of the income distribution would be relative to a given previous year if the tax schedule were augmented in order to hold changes in income inequality constant over time. Let $Q_0(\alpha)$ denote the α-quantile of the 2012 income distribution; let $Q_t(\alpha)$ denote the α-quantile of an alternative income distribution in year t. I define the inequality

[52] For example, re-ordering high income individuals with low-income individuals within the status quo environment would generate positive inequality deflated surplus (but would not change the value for most social welfare functions). This is primarily because the weights are higher for the low-income individuals and these weights do not change in the conceptual experiment.

[53] The data is constructed using Table 7 from CBO publication 43373. I take market income minus federal taxes to construct after-tax income shares across the population. To account for the fact that government spending may have value, I assign net tax collection back to each household in proportion to their after-tax income. This assumes each individuals' willingness to pay for government expenditure is proportional to after-tax income. The CBO also reports an "after-tax" measure of income that includes government transfers. Unfortunately, the bottom portion of the income distribution for these transfers disproportionately falls on the non-working elderly, through social security and Medicare payments. Since these would be affected by modifications to the nonlinear income tax schedule, I do not use this measure of income. I also do not make adjustments for local taxes (e.g. property taxes), since a spatial equilibrium model would predict these taxes would be associated with greater public good benefits and local amenities. Of course, future work could conduct the present analysis using a different surplus function. For example, in Appendix D, I conduct a similar analysis using the pre-tax income distribution using Census data going back to 1967. This would be a valid measure of the social cost of increased income inequality if the value of non-market goods were proportional to the tax revenue one paid. This is a potentially plausible model, but the natural benchmark is the case in which the non-market goods are valued proportional to after tax income.

[54] I construct the after-tax income series by taking market income minus federal taxes. I then add the difference between aggregate market income and after tax income (i.e. government spending/production) by assuming it is valued proportional to after-tax income; results are generally similar with other assumptions of the value of public expenditure across the income distribution.

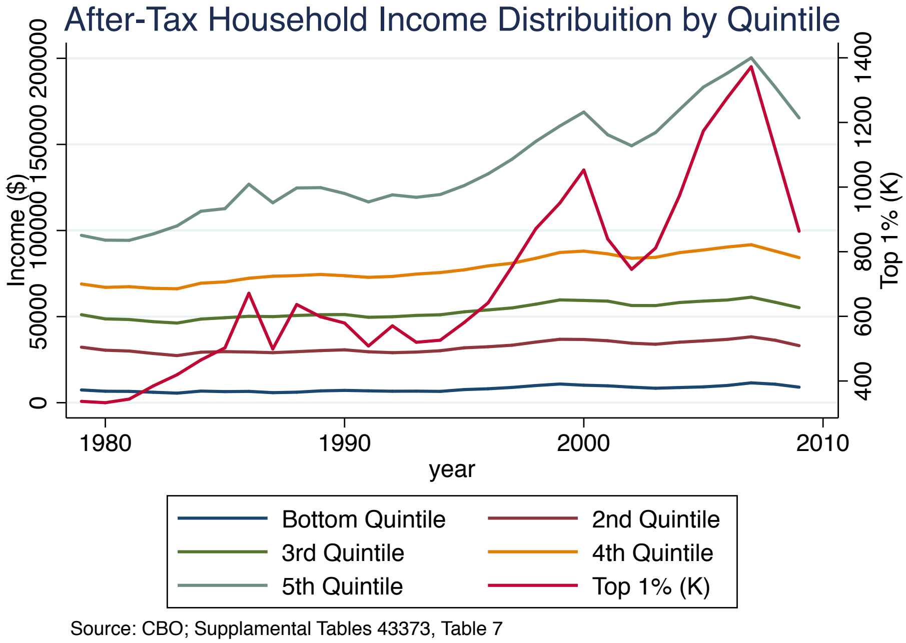

Figure 11: Figure presents the evolution of the distribution of after-tax income in the U.S. from 1979-2009 using CBO data.

deflated difference in household income by

$$S_t^{ID} = \int_0^1 \left[Q_0(\alpha) - Q_t(\alpha)\right] g^H\left(Q_0(\alpha)\right) d\alpha \tag{9}$$

where $g^H(y)$ is the inequality deflator constructed in the status quo. Intuitively, S_t^{ID} is the first-order approximation to the amount by which the U.S. would be richer in 2012 relative to year t if the 2012 income tax schedule were augmented in to hold constant the changes to the income distribution relative to year t.

To construct an estimate of inequality deflated growth in 2012 relative to previous years, I use the CBO data from 1979-2009, appended with Census data from 2009-2012.[55] All incomes are in units of 2012 income using the CPI-U deflator.[56]

Figure 12 reports the change in mean household income (dashed blue line), along with the inequal-

[55] I translate the CBO data from 2009 dollars to 2012 dollars using the CPI adjustment of 1/1.07. To fill in the years 2009-2012, I use IPUMS CPS micro-data to construct after-tax income measures analogous to the CBO data. I take total household income including government cash transfers and EITC, minus federal and state taxes. In practice, the Census data from 2009-2012 exhibits only minimal trends, and thus the social cost of income inequality from 1980-2012 is almost identical to that of 1980-2009. However, one potential bias is that Census data under-states income accumulation in the upper tail of the distribution (Piketty and Saez (2003)) and perhaps the recent recovery has been concentrated at upper regions of the income distribution (e.g. because of the stock market recovery). To this extent, the size of the GDP deflation required to account for increased income inequality is an under-estimate.

[56] This ignores potential changes in prices for goods with varying shares across the income distribution. An interesting and important direction for future work would be to construct a price deflator that uses consumption shares conditional on income, then aggregates across incomes using the inequality deflator.

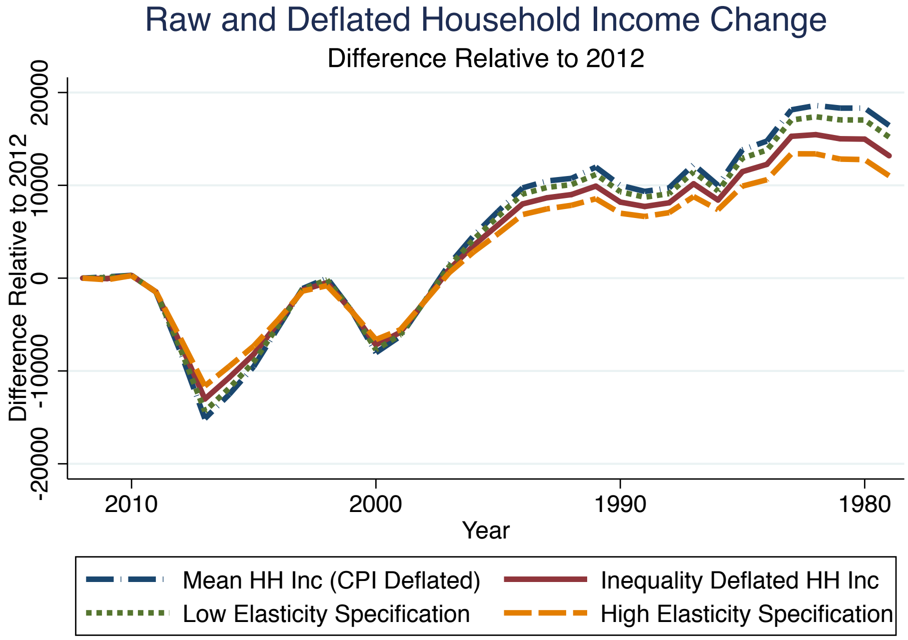

Figure 12: Raw and Deflated Household Income Change Relative to 2009 (All Income deflated to 2009 dollars using the CPI)

ity deflated change in household income under the baseline specification and two alternative elasticity specifications. Mean household income has increased by roughly \$18,300 relative to 1980, but inequality deflated growth has increased \$15,000 under the baseline specification (\$13K and \$17K under the high and low elasticity specifications, respectively). From a normative perspective, this lowers the overall growth rate of the U.S. economy by roughly 15-20%: if the U.S. were to make a tax adjustment so that everyone shared equally in the after tax earnings increases, roughly 15-20% of the growth since 1980 would be evaporated.

Figure 13 provides an estimate of the social cost of increased income inequality. To do so, I multiply the per-household social cost by the total number of households in the U.S.[57] This suggests the social cost of increased income inequality since 1980 is roughly \$400B. From an equivalent variation perspective, undoing the increased inequality would cost roughly \$400B; from a compensating variation perspective, if the U.S. had not experienced the increased inequality, it could have replicated the social surplus provided by the 2012 after tax income distribution even if aggregate economic growth were \$400B less than actually occurred.[58] Of course, these numbers depend on the behavioral responses to taxation – if one believes behavioral responses to taxes are larger (e.g. a compensated elasticity of 0.5), then the social cost of increased income inequality is in excess of \$600B.

[57]The census reports 117.6K households in 2009, with an annual increase over the years 2006-2009 of roughly 500 households per year, implying roughly 119K households in 2012.

[58]As discussed above, the compensating variation interpretation requires the additional assumption that the inequality deflator measures the marginal cost of taxation in the alternative environment.

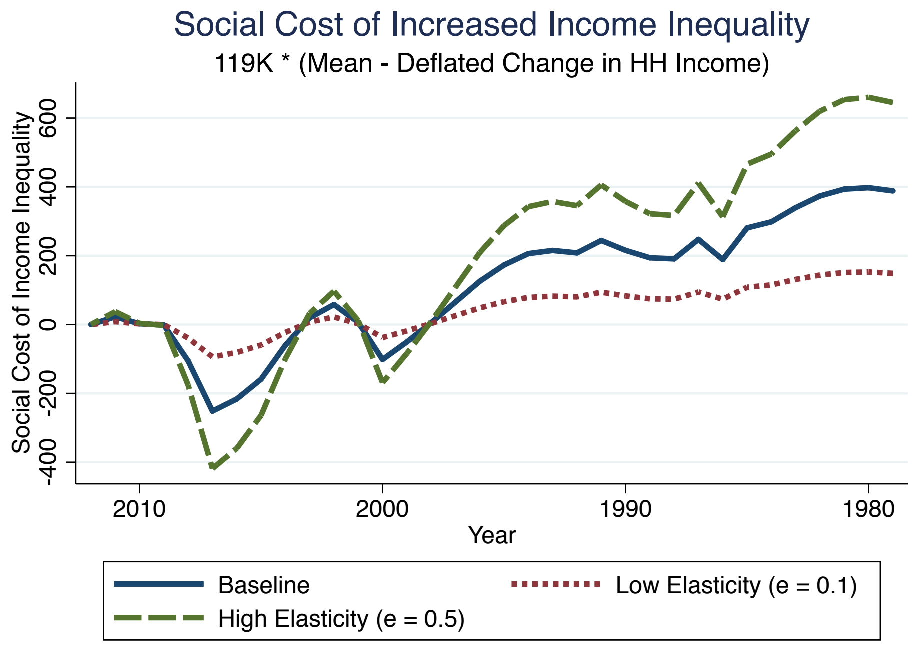

Figure 13: Social Cost of Increased Income Inequality

5.2 Comparisons of Income Distributions: Cross-Country Analysis

It is often noted that the U.S. has a higher degree of income inequality than many other countries of similar income per capita levels. This section uses the inequality deflator to ask how much richer or poorer the U.S. would be relative to these countries if it attempted to replicate their income distributions.

The inequality deflated surplus associated with the income distribution in country a is given by

$$S_a^{ID} = \int_0^1 \left[Q_a\left(\alpha\right) - Q_0\left(\alpha\right)\right] g^H\left(Q_0\left(\alpha\right)\right) d\alpha \tag{10}$$

To construct an estimate of inequality deflated surplus of each country's income distribution relative to the U.S., I use data from the World Bank Development Indicators and UN World Income Inequality Database. These sources aggregate household survey data from various countries and to provide measures of the shape of the income distribution. Appendix C discusses the data and estimation details.

Figure 14 plots deflated surplus against the GNI per capita of each country within $10,000 of the U.S. GNI per capita. The dots represent the estimates for the baseline specification and the brackets plot the estimates for the low and high elasticity specifications.

The results suggest that a couple of cross-country comparisons based on mean incomes are reversed when using the inequality deflator to control for differences in inequality. The U.S. is richer in mean per capita terms than Austria (AUT) and New Zealand (NLD) by roughly $2,000. However, the inequality

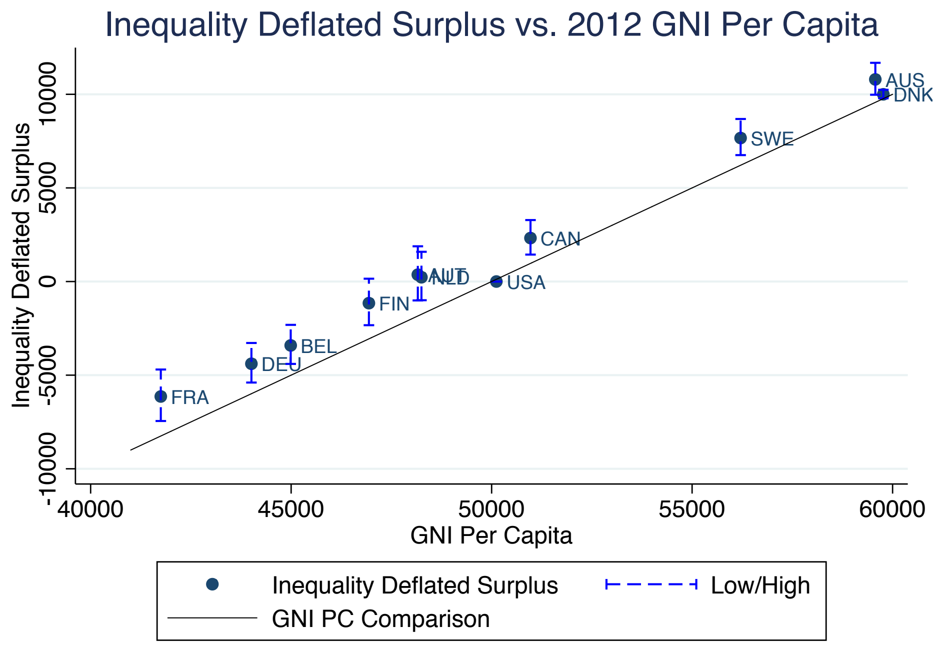

Figure 14: For each country, the inequality deflated surplus (defined in equation 10) is presented for the baseline elasticity specification against GNI per capita on the horizontal axis; vertical bars representing the high and low elasticity specifications. If all countries had the same degree of inequality (i.e. same deflation factors in Figure 18), then all countries would align on the 45 degree line. The fact that other countries lie above this 45 degree line reflects the greater degree of income inequality in the U.S. relative to these countries.

deflated surplus is positive for both countries ($227 for the Netherlands and $366 for Austria). Despite it's higher income level, if the U.S. were to try to provide the distribution of purchasing power offered by these countries, each point of the income distribution would be made worse off relative to these countries under the baseline elasticity specification. Under the high elasticity specification, Finland's income distribution has positive inequality deflated surplus, even though it has $3,180 less in per capita national income.

However, the broad pattern in the data suggest that the difference between inequality deflated surplus and total surplus does not completely change the ordering of countries. For example, the U.S. remains richer than Belgium, Germany, and France (in terms of inequality-deflated surplus) even for the high elasticity specification. So, adjusting for differences in income inequality leads to some reversals in the ordering of income distributions, but the results suggest that comparisons generally follow fairly closely to mean income per capita comparisons.[59]

[59]This is similar to findings of Jones and Klenow (2011) who suggest differences in measures of ex-ante welfare track GDP fairly well.

6 Welfare Analysis of Public Expenditure

The previous section illustrates how a dollar of surplus that falls to the poor is more valuable than a dollar that falls to the rich. This section explores the policy implications. On the one hand, the inequality deflator suggests a bias towards "pro poor" policies. But, when considering policy experiments, there is an important counteracting force: policies that conduct redistribution may have similar fiscal externalities to those induced by changes in the income tax schedule.

This section walks through two conceptualization of policy experiments, and provides examples of how the inequality deflator can be used to make normative policy statements based on the potential Pareto principle. First, one can imagine a budget neutral policy experiment that delivers surplus, $s(\theta)$, to each person θ. Second, one can imagine a non-budget neutral policy experiment that delivers surplus, $s(\theta)$. I begin with the case of budget neutral policy experiments.

6.1 Budget Neutral Policies and Producer vs. Consumer Surplus

If $s(\theta)$ corresponds to a budget neutral policy experiment, then measuring inequality deflated surplus characterizes the existence of potential Pareto improvements (where if there is heterogeneity conditional on income, one needs to use max/min surplus). Of course, in many practical examples, the groups of people affected will often be aggregated. For example, perhaps an economist has estimates of the impact of a merger on surplus across the income distribution. Moreover, suppose producer surplus falls solely on the top 1% and consumer surplus falls evenly across the income distribution. The inequality deflator would suggest one should weight producer surplus at 0.77 relative to consumer surplus at 1:

$$S^{ID} = 0.77S^P + S^C$$

It is commonplace to apply a "consumer surplus standard" to merger policies (e.g. Nocke and Whinston (2014)) on distributional grounds, providing no weight to consumer surplus. The inequality deflator suggests one should give substantial weight to producer surplus regardless of one's social preferences, even if all producer surplus falls to the top 1%.

Of course, this is a weighting that applies to the U.S. regulatory policy. In other countries, the stringency of merger approvals should depends on their cost of redistribution. For example, some have argued that the top tax rate in France is at or above the revenue-maximizing level (Bourguignon and Spadaro (2012)), which suggests the French inequality deflator is close to zero in the upper regions of the income distribution. In France, a consumer surplus standard for merger approval may be optimal.

Finally, it cannot be understated that the preceding analysis relies heavily on the assumption that the economist has constructed surplus, $s(\theta)$, from a budget neutral policy. For example, approving mergers that increase incomes of the rich may cause them to increase or decrease their taxable income – these effects would need to be incorporated into the analysis. For these cases, it will be useful to have a method of analysis for non-budget neutral policies. It is to this I now turn.

6.2 Non-Budget Neutral Policies

Now, consider a policy that is non-budget neutral. In particular, imagine that the policy increases government spending by \$1 on G. G could be any item of the government's budget, including public goods (e.g. roads, public transit, schools), R&D subsidies (e.g. tax credit to multi-national companies, SBA loans that provide credit to those with disadvantaged status, etc.), targeted conditional transfers (e.g. food stamps, housing subsidies, mortgage interest deductions, etc.), education vouchers, or other taxes/subsidies (e.g. capital taxes, commodity taxes, corporate taxes, etc.).

6.2.1 Searching for Pareto Improvements

Given the policy, each individual, θ, has a marginal willingness to pay, $s(\theta)$.[60] In practice, estimation of this willingness to pay could be quite difficult. For example, an individual's marginal willingness to pay for a change in the corporate tax policy depends upon the precise incidence of corporate taxes. For the present purposes, I imagine the economist has constructed these measures of individual willingness to pay; and I consider the question of aggregation across individuals.

To do so, I construct a generalized "benefits principle" (Hylland and Zeckhauser (1979), Kaplow (1996, 2004, 2006, 2008); see Kreiner and Verdelin (2012) for a summary). To search for potential Pareto improvements, one can construct the minimum and maximum surplus functions from above,

$$\overline{s}(y) = \sup\{s(\theta)\,|y(\theta) = y\}$$

$$\underline{s}(y) = \inf\{s(\theta)\,|y(\theta) = y\}$$

In addition to each individual's willingness to pay, the additional spending has a total marginal cost given by

$$Cost = 1 + FE^G$$

which includes not only the \$1 mechanical increase in the cost of G, but also the aggregate fiscal externality, FE^G, per dollar of mechanical government spending.

When can additional spending on G result in a potential Pareto improvement? Imagine the government provides transfers so that everyone with income y obtains $\underline{s}(y)$ of benefits. Doing so will be feasible if and only if

$$\underbrace{\int (1 + FE(y))\,\underline{s}(y)\,dF(y)}_{\text{Taxed Benefits}} > \underbrace{1 + FE^G}_{\text{Fiscal Costs}} \tag{11}$$

where $F(y)$ is the distribution of income in the status quo world. If equation (11) holds, the government can tax all the beneficiaries and compensate the losers so that everyone is better off from additional spending on G. Conversely, when does less spending on G result in a potential Pareto improvement? Now, the government must compensate $-\overline{s}(y)$ at each income level and the policy has a marginal cost

[60] For example, if individuals have utility functions $u(c, y, G; \theta)$ and the policy changes G by $dG(\theta)$ for each type, then $s(\theta) = \frac{\frac{\partial u}{\partial G}}{\frac{\partial u}{\partial c}} dG(\theta)$.

of $-\left(1+FE^{G}\right)$. So, a potential Pareto improvement from less G occurs if and only if

$$\underbrace{\int \left(1+FE\left(y\right)\right)\overline{s}\left(y\right)dF\left(y\right)}_{\text{Taxed Benefits}} < \underbrace{1+FE^{G}}_{\text{Fiscal Costs}} \tag{12}$$

Equations (11) and (12) characterize when there exists modifications to non-budget neutral government policies that can be combined with modifications to the tax schedule to deliver a Pareto improvement.[61]

Relation to Samuelson (1954) Equations (11) and (12) provide modified Samuelson condition (Samuelson (1954)) that accounts for the distributional incidence of the policy change without requiring a social welfare function. Instead, one needs to construct the inequality-deflated surplus (or min/max surplus), which requires knowledge of individuals' willingness to pay for the policy change, $s\left(\theta\right)$, and the aggregate fiscal externality of increasing G, FE^{G}. The fiscal externality component does not require a decomposition of behavioral responses into income and substitution effects (Hendren (2013)). As a result, these equations provide a welfare framework that not only does not require a social welfare function, but also only requires estimates of the willingness to pay for the policy and the causal effects of the policy, both of which are generally the focus of empirical work.

Implicitly, equations (11) and (12) are comparing the increased spending on G to the distortionary impact of the tax schedule. Indeed, it is straightforward to verify that if G is a modification to the ordinary income tax schedule, then $\int \underline{s}\left(y\right)dF\left(y\right) = \int \overline{s}\left(y\right)dF\left(y\right) = 1$ and $FE^{G} = \int FE\left(y\right)s\left(y\right)$, so that both equations (11) and (12) hold with equality. Intuitively, the potential Pareto criteria leads to indifference over the shape of the current income tax schedule.[62]

Remark 1. *(Indifference to changes in $T\left(y\right)$) Unless there are local Laffer effects (so that $g\left(y\right)<0$ and $FE\left(y\right)<0$), modifications to the income tax schedule cannot generate potential Pareto improvements.*

Relation to Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2008) More generally, whether one prefers policies that provide benefits to the poor or to the rich depends on the fiscal externality of the policy change, FE^{G}. One special case that has been noted in the literature is when spending on G has the same fiscal externality impact per unit of welfare provided as does changes in the income tax. In this case, one recovers a traditional Samuelson condition.

Remark 2. (Un-weighted Samuelson condition) *Suppose increased spending on G has a distortionary impact on tax revenue similar to changes in the income tax schedule (i.e. $FE^{G} = \int s\left(\theta\right)FE\left(y\left(\theta\right)\right)d\mu\left(\theta\right)$). Then, if benefits do not vary conditional on income, additional spending on G generates potential*

[61]Note that the LHS of these equations is very close to inequality deflated max and min surplus, but differ to the extent to which $E\left[1+FE\left(y\right)\right]\neq 1$ (i.e. there are income effects so that the aggregate fiscal externality of a lump-sum tax is not 1). Intuitively, these equations are making welfare comparisons using dollars in the hand of the government; whereas the inequality deflator divides by $E\left[1+FE\left(y\right)\right]$ so that it corresponds to an experiment that holds money in the hands of the population. In practice, I estimate $E\left[1+FE\left(y\right)\right]\approx 0.98$, so that one can empirically think of the LHS of these equations as max and min deflated surplus.

[62]Unless $FE\left(y\right)<-1$, so that modifications to the existing tax schedule led to a Pareto improvement.

Pareto improvements if and only if

$$\int s(\theta)\,d\mu(\theta) \geq 1$$

If surplus varies conditional on income, then additional spending on G generates potential Pareto improvements if and only if $\int \underline{s}(\theta)\,d\mu(\theta) \geq 1$ and tax schedule modifications in the status quo generate a potential Pareto improvement relative to spending on G if and only if $\int \overline{s}(\theta)\,d\mu(\theta) \geq 1$. In all cases, one weights surplus equally on the rich and the poor.

If spending on G has the same distortionary impacts on taxable behavior as does changes to the nonlinear income tax schedule, then whether one wishes to provide additional G depends on the aggregate (unweighted) willingness to pay for the public good relative to its cost (also known as a "standard" Samuelson condition). This result has been derived in existing theoretical work (Hylland and Zeckhauser (1979); Kaplow (1996, 2004, 2008)) under the special case in which G and c are weakly separable in the utility function from y.[63] If this particular form of weak separability holds for the utility function, then Remark 2 generalizes this result to show that one only require that, on average, the fiscal externality induced by changes in G is similar to the fiscal externality of an analogous change in the tax schedule. If G has the same distortionary properties as changes in the income tax schedule, individuals' (unweighted) aggregate willingness to pay characterizes whether additional spending is desirable.

Importantly, the cost-benefit equations ((11) and (12)) provide a generalization of the benefits principle to the case where G is not weakly separable and the fiscal externalities associated with the provision of G are systematically different from changes in the tax schedule. Fortunately, these cases are handled by simply comparing the inequality-deflated benefits to the aggregate costs, inclusive of the fiscal externality. For example, the empirical literature on the provision of in-kind transfers often finds minimal impacts of such transfers on labor supply. Often, this is interpreted as suggesting that in-kind redistribution may not be superior to redistribution through the tax schedule (Currie and Gahvari (2008)). But, the the absence of a negative fiscal externality from in-kind transfers suggests a rationale for redistribution through in-kind transfers, as opposed to redistribution through the tax schedule, which entails a fiscal externality. If $FE^G = 0$, then clearly one wishes to weight the surplus to the poor more so than to the rich.

Remark 3. *(Tagging/screening) The social value of targeted publicly provided goods which do not induce behavioral responses that affect taxable income[64] is higher for the poor relative to the rich.*

6.2.2 Targeted Non-Budget Neutral Policies

In many cases, government policies are targeted at fairly specific regions of the income distribution. In these cases, I illustrate how the analysis can be simplified even further if one is willing to make a

[63] Under weak separability, one can show that $FE^G = \int FE^G(y(\theta))\,d\mu(\theta)$ where $FE^G(y) = E[s(\theta)\,FE(y(\theta))\,|y(\theta) = y]$. Remark 2 shows that this condition need not hold for all y, but rather can hold only on average.

[64] *For example, this would be true if the policy affects goods which are additively separable in the utility function, $\psi(P) + u(c, y; \theta)$ so that P does not affect labor supply.*

couple of additional assumptions about who benefits from the policy change. Suppose the increased spending of \$1 on G is targeted solely to those earning near a given income amount, y. Let $s(y)$ denote these individual's willingness to pay for the additional spending on G, and assume that this is homogeneous conditional on income.

Now, let $1 + FE^G$ denote the aggregate cost of spending these resources on G. One can construct the marginal value of public funds (MVPF as in Hendren (2013)[65]) of additional spending on G as the individuals' willingness to pay for G divided by its cost

$$MVPF_G = \frac{s(y)}{1 + FE^G}$$

This is a simple cost-benefit measure of the value of the additional government spending.

Given the MVPF, one can compare it to the cost of providing the same welfare benefits to these individuals, except through modifications to the income tax schedule:

$$MVPF^{Tax}(y) = \frac{1}{1 + FE(y)}$$

These individuals value the transfer at \$1 and it has a cost of $1 + FE(y)$, where $FE(y)$ is the fiscal externality associated with providing the additional transfer to those earning y. Given these definitions, additional spending on the targeted policy G yields a Pareto improvement if and only if

$$\frac{s(y)}{1 + FE^G} > \frac{1}{1 + FE(y)}$$

Examples Figure 15 provides an illustration of this approach using the MVPF for several policies, taken from the calculations in Table 1 of Hendren (2013). In particular, I include the MVPF from Section 8 housing policy, drawing upon estimates of the fiscal externality from the causal effects in Jacob and Ludwig (2012); Food stamps drawing upon estimates of the labor supply fiscal externalities from Hoynes and Schanzenbach (2012); and the job training partnership program, drawing upon estimates from Bloom et al. (1997).

For each of these policies, I take their implied MVPF (from the calculations in Hendren (2013)) and plot them at the mean income of their program participants.[66] The results suggest that policies such as food stamps and housing vouchers may be less efficient forms of redistribution than modifications to the income tax schedule (i.e. EITC for these beneficiaries). Also, the JTPA appears to have a significantly positive welfare impact, which is driven by its positive fiscal externalities discussed in Hendren (2013).[67]

[65]This measure of the MVPF was originally defined in Mayshar (1990), and subsequently used by Slemrod and Yitzhaki (2001) and Kleven and Kreiner (2006).

[66]The estimate of the MVPF for food stamps comes from the introduction of food stamps in the 1960s and 70s. I translate the beneficiaries' mean incomes in the samples of Hoynes and Schanzenbach (2012) to 2012 dollars. But, one should keep in mind that the food stamp program today affects a larger fraction of the population and reaches a further up the income distribution.

[67]However, an alternative interpretation is that the beneficiaries of these policies are different than the beneficiaries of modifications to the tax schedule. For example, food stamps may target those less able to work, conditional on income,

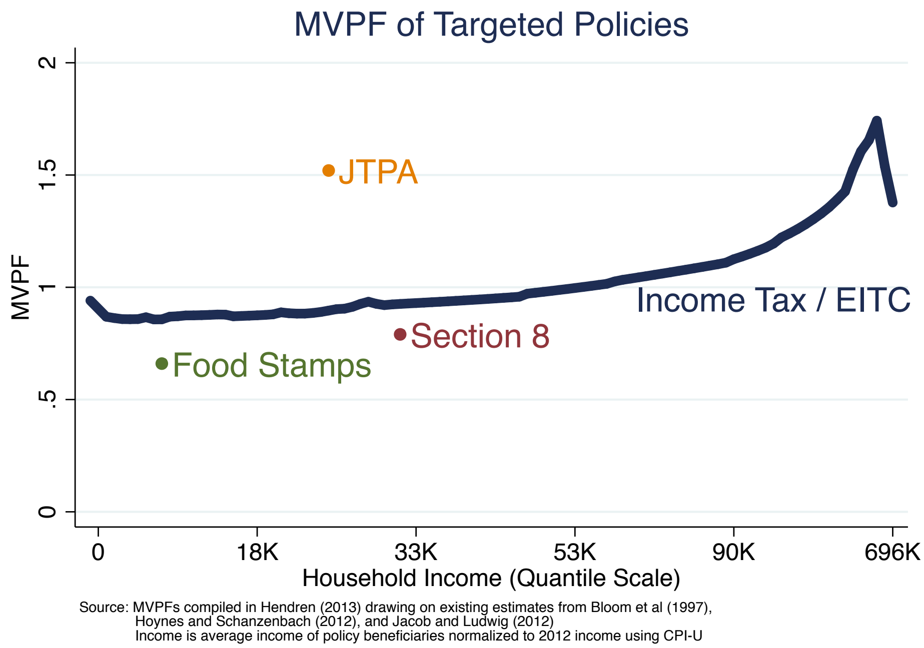

Figure 15: Figure presents values of the marginal value of public funds (MVPF) on the vertical axis and income quantiles on the horizontal axis. MVPFs for the job training partnership act (JTPA), food stamps, and section 8 housing vouchers are constructed using existing estimates of individual willingness to pay and causal effects of these policies on the government budget. See Table 1 of Hendren (2013) for details). For each policy, the MVPF is plotted on the horizontal axis at the mean income of the beneficiaries of the policy. Assuming no heterogeneity in benefits of policies conditional on income, MVPFs falling above (below) the the Income/EITC line correspond to policies that can (not) generate Pareto improvements.

Although these estimates are taken from existing literature, they are of course subject to various estimation and identification concerns discussed in Hendren (2013). So one should use caution before making precise policy prescription.[68] But, the key limitation to providing policy prescriptions is the quality of the causal effects and measures of individual willingness to pay for policy changes, not the need to resolve interpersonal comparisons or decompose behavioral responses into income and substitution effects (which would be required for computing, e.g., a measure of the "deadweight loss" associated with these policy changes).[69]

If the willingness to pay for the policy change is not heterogeneous conditional on income, then additional spending on G delivers a potential Pareto improvement iff it has a higher MPVF than the

relative to the EITC.

[68]For example, one cannot reject that the MVPF for Food Stamps and Section 8 equals 1; also, there is considerable uncertainty about individuals' willingness to pay for the JTPA. The MVPF for the JTPA in Figure 15 assumes individuals valued the program at its cost. But, this would understate the MVPF if individuals valued the benefits at their observed earnings increases; it would over-state the MVPF if individuals did not value the program's benefits (which is plausible because it was a heavily subsidized program). See Hendren (2013) for a further discussion.

[69]See Hendren (2013) for a further discussion of the difference between the MVPF and deadweight loss or the traditional measures of the marginal cost of public funds (MCPF).

corresponding MVPF for transfers through the tax schedule, as illustrated in Figure 15. Conversely, less spending on G delivers a potential Pareto improvement if and only if it has a lower MVPF than the corresponding MVPF for the tax schedule.[70]

While the canonical Kaldor-Hicks compensation principle compares policies to the efficiency of lump-sum transfers, Figure 15 provides a graphical method for comparing policies to the efficiency of modifications to the income tax schedule. This implements a welfare framework based on feasible redistribution through the tax schedule. The resulting recommendations do not rest on the subjective choices made by the researcher to resolve interpersonal comparisons. Rather, they rest on the capability of the researcher to estimate measures of individual willingness to pay for the policy and the causal effect of the policy on the government's budget.

7 Conclusion

The inequality deflator provides a method for resolving interpersonal comparisons that does not require a social welfare function. Building on the Kaldor-Hicks compensation principle, it provides a method for comparing environments using the Pareto criteria combined with hypothetical modifications to the nonlinear income tax schedule. Existing empirical evidence suggests distortions in the tax system make transfers from rich to poor roughly twice as costly as transfers from poor to rich. Therefore, the inequality deflator weights surplus accruing to the poor roughly twice as much as surplus accruing to the rich to account for the cost of spreading these benefits equally across the income distribution.

There are many potential directions for future work. One could extend the set of feasible transfers beyond the income tax schedule to also include other policies (e.g. augmentations to both the income tax and the generosity of social safety net programs such as Medicaid). Such an approach may provide better guidance in cases where surplus varies conditional on income and provide tighter bounds for searches for Pareto improvements in cases with heterogeneity conditional on income. Further analysis could also incorporate other constraints to feasible redistribution, such as political economy constraints, and attempt to incorporate these additional shadow prices into the cost of redistribution. From an econometric perspective, the present analysis highlights the importance of estimating not only the average willingness to pay for policy changes, but also its weighted average across the income distribution (i.e. inequality deflated surplus), and properties of its minimum and maximum conditional on income.

By relying on the potential Pareto criterion, the inequality deflator provides a welfare criteria that avoids the inherent subjectivity of the social welfare function approach. In the end, there will always be disagreement about the extent to which society should redistribute from those with high to those

[70] If willingness to pay is heterogeneous conditional on income, then one can construct minimum and maximum MVPFs using $\underline{s}(y)$ and $\overline{s}(y)$. Alternatively, if the researcher cannot estimate such heterogeneity in the willingness to pay but suspects it may be an important component of the policy analysis, then the degree to which the MVPF falls below the tax schedule reveals the implicit cost of providing benefits to individuals who benefit from G as opposed to providing transfers. Hence, one can have a back-of-the-envelope estimate of the size of the "covariance bias" term in Corollary 2, which reveals how society is implicitly willing to transfer resources to beneficiaries of G relative to an income transfer to those earning y.

with low incomes. But, the inequality deflator can help generate broader agreement on the desirability of a wide range of other policies that require a resolution of interpersonal comparisons.

References

Atkinson, A. B. and N. H. Stern (1974, January). Pigou, taxation and public goods. *The Review of Economic Studies 41*(1), 119–128. 6

Atkinson, A. B. and J. E. Stiglitz (1976). The design of the tax structure: Direct versus indirect taxation. *Journal of Public Economics 6*(1-2). 5, 3.2, 30

Ballard, C. L. and D. Fullerton (1992). Distortionary taxes and the provision of public goods. *Journal of Economic Perspectives 6*(3), 117–131. 6

Bargain, O., M. Dolls, D. Neumann, A. Peichl, and S. Siegloch (2011). Tax-benefit systems in europe and the us: between equity and efficiency. *CESifo Working Paper No. 3534*. 1, 3.3, 3.6

Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics 52*(2), 310–34. 1

Bloom, H., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Boss (1997). The benefits and costs of jtpa title ii-a programs: Key findings from the national job training partnership act study. *The Journal of Human Resources 32*(3), 549–576. 6.2.2

Blundell, R., M. Brewer, P. Haan, and A. Shephard (2009). Optimal income taxation of lone mothers: An empirical comparison of the uk and germany. *The Economic Journal 119*, 101–121. 1, 3.3, 3.6

Boadway, R. (1974). The welfare foundations of cost-benefit analysis. *The Economic Journal 84*(336), 926–939. 1, 5

Bourguignon, F. and A. Spadaro (2012). Tax-benefit revealed social preferences. *Journal of Economic Inequality 10*, 75–108. 1, 8, 3.3, 3.6, 48, 6.1

Browning, E. and W. R. Johnson (1984). The trade-off between equality and efficiency. *The Journal of Political Economy 92*(2), 175–203. 34

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance. *American Economic Journal: Economic Policy 1*(2), 31–52. 3.8

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018. 1, 4.2

Chetty, R., J. Friedman, and E. Saez (2013). Using differences in knowledge across neighborhoods to uncover the impacts of eitc on earnings. *American Economic Review (Forthcoming)*. 1, 3.3, 4.2

Chetty, R., N. Hendren, P. Kline, and E. Saez (2013). Where is the land of opportunity: The geography of intergenerational mobility in the united states. *Working Paper*. 4.1, 4.1

Coate, S. (2000). An efficiency approach to the evaluation of policy changes. *Economic Journal 110*(463), 437–455. 1, 4

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume Volume 1, pp. 87–166. Elsevier. 49

Currie, J. and F. Gahvari (2008). Transfers in cash and in-kind: Theory meets the data. *Journal of Economic Literature 46*(2), 333–83. 6.2.1

Diamond, P. A. and J. A. Mirrlees (1971, June). Optimal taxation and public production II: tax rules. *The American Economic Review 61*(3), 261–278. 1

Diamond, P. A. and E. Saez (2011). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives 25*(4), 165–90. 38, 4.3, B

Dreze, J. and N. Stern (1987). *Handbook of Public Economics*, Chapter The Theory of Cost-Benefit Analysis. Elsevier Science Publishers. 1, 4, 1

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics 81*(4), 674–680. 3.8

Fleurbaey, M. (2009). Beyond gdp: The quest for a measure of social welfare. *Journal of Economic Literature 47*(4), 1029–1075. 1

Gruber, J. and E. Saez (2002, April). The elasticity of taxable income: evidence and implications. *Journal of Public Economics 84*(1), 1–32. 4.2

Harberger, A. (1971). Three basic postulates for applied welfare economics: An interpretive essay. *Journal of Economic Literature 9*(3), 785–797. 2

Hendren, N. (2013). The policy elasticity. *NBER Working Paper No. 19177*. 1, 6, 1, 20, 3.3, 45, 6.2.1, 6.2.2, 6.2.2, 15, 68, 69

Hicks, J. R. (1939). The foundations of welfare economics. *Economic Journal 49*, 696–712. 1, 3, 3.5

Hicks, J. R. (1940). The valuation of social income. *Economica 7*(26), 105–124. 1, 26, 3.5

Hotz, J. and K. Scholz (2003). The earned income tax credit. In R. A. Moffit (Ed.), *Tax Policy and the Economy*. University of Chicago Press. 1, 3.3, 45

Hoynes, H. W. and D. W. Schanzenbach (2012). Work incentives and the food stamp program. *Journal of Public Economics 96*(1-2), 151–162. 6.2.2, 66

Hylland, A. and R. Zeckhauser (1979). The efficient allocation of individuals to positions. *The Journal of Political Economy*, 293–314. 1, 16, 6.2.1, 6.2.1

Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez (2007). Welfare reform in european countries: A microsimulation analysis. *The Economic Journal 117*, 1–44. 34

Jacob, B. A. and J. Ludwig (2012). The effects of housing assistance on labor supply: Evidence from a voucher lottery. *American Economic Review 102*(1), 272–304. 6.2.2

Jones, C. I. and P. J. Klenow (2011). Beyond gdp? welfare across countries and time. *Working Paper*. 12, 59

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal 49*(195), 549–552. 1, 3.5, 5

Kaplow, L. (1996). The optimal supply of public goods and the distortionary cost of taxation. *National Tax Journal 49*, 513–534. 1, 16, 6.2.1, 6.2.1

Kaplow, L. (2004). On the (ir)relevance of the distribution and labor supply distortion to government policy. *Journal of Economic Perspectives 18*, 159–175. 1, 16, 6.2.1, 6.2.1

Kaplow, L. (2006). On the undesirability of commodity taxation even when income taxation is not optimal. *Journal of Public Economics 90*, 1235–1250. 16, 6.2.1

Kaplow, L. (2008). *The Theory of Taxation and Public Economics*. Princeton University Press. 1, 16, 6.2.1, 6.2.1

Kleven, H. J. and C. T. Kreiner (2006). The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics*, 1955–1973. 6, 34, 35, 65

Kreiner, C. T. and N. Verdelin (2012). Optimal provision of public goods: A synthesis. *Scandinavian Journal of Economics 114*(2), 384–408. 16, 17, 6.2.1

Liebman, J. and E. Saez (2006). Earnings responses to increases in payroll taxes. *Working Paper*. 4.2

Lockwood, B. B. and M. Weinzierl (2014). The evolution of revealed social preferences in the united states and the costs of unequal growth and recessions. *Working Paper*. 1

Mayshar, J. (1990). On measures of excess burden and their applications. *Journal of Public Economics 43*(3), 263–89. 65

Mirrlees, J. (1976, November). Optimal tax theory: A synthesis. *Journal of Public Economics 6*(4), 327–358. 1

Mirrlees, J. A. (1971). An exploration into the theory of optimal income taxation. *The Review of Economic Studies 38*(2), 175–208. 1, 3, 3.2

Nocke, V. and M. Whinston ((2014)). Merger policy with merger choice. *American Economic Review (Forthcoming).* 6.1

Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913-1998. *Quarterly Journal of Economics 118*(1), 1–41. 1, 5.1, 55, D

Piketty, T. and E. Saez (2007). *Top Incomes over the Twentieth Century. A Contrast Between Continental European and English-Speaking Countries*, Chapter Income and Wage Inequality in the United States 1913-2002. Oxford University Press. D, 82

Piketty, T. and E. Saez (2012). Optimal labor income taxation. *NBER Working Paper No. 18521*. 38, 4.3, B

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the roy model. *Quarterly Journal of Economics 128*, 623–668. 3.8

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*, 205–229. 1, 3, 3.3, 3.3, 47, B, 74

Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*(3), 1039–1073. 3, 3.6

Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50. 1, 3.3, 3.8, 4.1, 4.2

Saez, E. and S. Stantcheva (2013). Generalized social marginal welfare weights for optimal tax theory. *NBER Working Paper No. 18835*. 3.6

Samuelson, P. (1947). *Foundations of Economic Analysis.* Harvard University Press. 1

Samuelson, P. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics 36*, 387–389. 6.2.1

Schlee, E. E. (2013). Radner's cost-benefit analysis in the small: An equivalence result. *Working Paper*. 26

Scitovsky, T. (1941). A note on welfare propositions in economics. *Review of Economic Studies 9*, 77–88. 5, 26

Slemrod, J. and S. Yitzhaki (2001). Integrating expenditure and tax decisions: The marginal cost of funds and the marginal benefit of projects. *National Tax Journal*. 1, 65

Stiglitz, J. E. and P. Dasgupta (1971, April). Differential taxation, public goods, and economic efficiency. *The Review of Economic Studies 38*(2), 151–174. 6

Werning, I. (2007). Pareto efficient income taxation. *MIT Working Paper*. 1, 9, 3.2, 46

Zoutman, F. T., B. Jacobs, and E. L. W. Jongen (2013a). Optimal redistributive taxes and redistributive preferences in the netherlands. *Mimeo*. 1, 8, 3.3, 3.6

Zoutman, F. T., B. Jacobs, and E. L. W. Jongen (2013b). Revealed social preferences of dutch political parties. *Working Paper*. 1, 8, 3.3, 38

Appendix (Not for Publication)

A Proofs

A.1 Proof of Proposition 1

Recall $\mu(\theta)$ is the probability over the type space Θ and $F(y)$ is the cumulative distribution of income in the status quo, $F(x) = \int 1\{y(\theta) \leq x\} d\mu(\theta)$. Recall I have defined $\hat{T}(y; y^*, \epsilon, \eta)$ by

$$\hat{T}(y; y^*, \epsilon, \eta) = \begin{cases} T(y) & if\ y \notin \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \\ T(y) - \eta & if\ y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \end{cases}$$

so that $\hat{T}$ provides η additional resources to an ϵ-region of individuals earning between $y^* - \epsilon/2$ and $y^* + \epsilon/2$. Fix ϵ and $\hat{q}(y^*, \epsilon, \eta)$ denote the net government resources expended under policy $\hat{T}(y; y^*, \epsilon, \eta)$ of the policy on the government's budget. Given the tax schedule, $\hat{T}(y; y^*, \epsilon, \eta)$, let $\hat{y}(\theta; y^*, \epsilon, \eta)$ denote the individual θ's choice of earnings, y. The net resources expended is given by:

$$\hat{q}(y^*, \epsilon, \eta) = \frac{-1}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} \int_\theta \hat{T}(\hat{y}(\theta; y^*, \epsilon, \eta); y^*, \epsilon, \eta) d\mu(\theta)$$

One needs to evaluate this derivative wrt η at $\eta = 0$.[71] WLOG, I can assume $\hat{q}(y^*, \epsilon, 0) = 0$.

A.1.1 Continuous Responses Only

To begin, I assume there is no participation margin response. Specifically, I assume that preferences are convex in consumption-earnings space so that $\hat{y}(\theta; y^*, \epsilon, \eta)$ is continuously differentiable in η. Below, I add back in extensive margin responses, so that $\hat{y}$ can move from 0 to interior earnings points in response to a change in η.

Of course, individuals may have different curvatures of their utility function. To capture this, define $c(y; \theta)$ to be the individual θ's indifference curve in consumption-earnings space at the baseline utility level. Given an agent θ's choice $y(\theta)$ facing the baseline tax schedule $T(y)$, the indifference curve solves

$$u(c(y; \theta), y; \theta) = u(T(y(\theta)) - y(\theta), y(\theta); \theta)$$

Note that the individual's first order condition requires:

$$c'(y(\theta); \theta) = -\frac{u_y}{u_c} = 1 - T'(y(\theta)) \tag{13}$$

[71] The main goal of the analysis will be to characterize the marginal cost of a policy. But, for relationship to the social welfare function, note that the envelope theorem suggests that the welfare impact of the policy is given by

$$\frac{d\hat{W}}{d\eta}|_{\eta=0} = \int 1\left\{y(\theta) \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)\right\} G'(\theta) dF(\theta)$$

where $g'(\theta)$ is the social marginal utility of income at θ and $dF(\theta)$ is the measure of individuals at θ.

so that the slope of this indifference curve equals the marginal keep rate, $1 - T'$.

In addition, the curvature of this indifference curve governs the size of the fraction of people who change their behavior in order to obtain the transfer, η. Let $k(\theta) = c''(y(\theta);\theta)$ denote the curvature of the indifference curve of type θ in the status quo world. First, consider those whose baseline income is just above $y^* + \frac{\epsilon}{2}$ but the opportunity to obtain the η transfer induces them to drop their income down to $y^* + \frac{\epsilon}{2}$. For individuals with curvature k, a second-order expansion of c (i.e. first order expansion of c') shows that anyone between $y^* + \frac{\epsilon}{2}$ and $y^* + \frac{\epsilon}{2} + \gamma(\eta; k)$ will choose incomes at $y^* + \frac{\epsilon}{2}$, where $\gamma(\eta; k)$ solves

$$\frac{(\gamma(\eta; k))^2}{2} k = \eta$$

or

$$\gamma(\eta; k) = \sqrt{\frac{2\eta}{k}}$$

Similarly, for individuals with curvature k, those with incomes between $y^* - \frac{\epsilon}{2} - \gamma(\eta; k)$ and $y^* - \frac{\epsilon}{2}$ will choose to increase their incomes to $y^* - \frac{\epsilon}{2}$.

Given these definitions, one can write the budget cost as the sum of four terms:

$$\int_\theta \hat{T}(\hat{y}(\theta; y^*, \epsilon, \eta); y^*, \epsilon, \eta)\, d\mu(\theta) = A + B + C + D + o(\eta)$$

where $\lim_{\eta \to 0} \frac{o(\eta)}{\eta} = 0$ (so that $\frac{do}{d\eta}|_{\eta=0} = 0$, so that one can ignore this term in the calculation of $\frac{d\hat{q}}{d\eta}|_{\eta=0}$).

The first term, A is the mechanical cost that must be paid to all those who receive the η transfer.

$$A = \eta \int 1\left\{ y(\theta) \in \left(y^* - \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k(\theta)}}, y^* + \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k(\theta)}} \right) \right\} d\mu(\theta)$$

The second term is the cost from those with baseline earnings above $y^* + \frac{\epsilon}{2}$ who drop their income down to $y^* + \frac{\epsilon}{2}$,

$$B = \int \left(T\left(y^* + \frac{\epsilon}{2}\right) - T(y(\theta)) \right) 1\left\{ y(\theta) \in \left(y^* + \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k(\theta)}} \right) \right\} d\mu(\theta)$$

And, conversely, the third term is from those with baseline earnings below $y^* - \frac{\epsilon}{2}$ who increase their incomes to $y^* - \frac{\epsilon}{2}$,

$$C = \tau\left(y - \frac{\epsilon}{2}\right) \int \left[y(\theta) - \left(y^* - \frac{\epsilon}{2}\right) \right] 1\left\{ y(\theta) \in \left(y^* - \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k(\theta)}}, y^* - \frac{\epsilon}{2} \right) \right\} d\mu(\theta)$$

and finally the fourth term is the income effect on earnings for those with baseline earnings in the

ϵ-region near y^*,

$$D = \int \left[T\left(\hat{y}\left(\theta; y^*, \epsilon, \eta\right)\right) - T\left(y\right)\right] 1\left\{y\left(\theta\right) \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)\right\} d\mu\left(\theta\right)$$

The remaining term, $o\left(\eta\right)$, captures the bias from approximating the B and C terms using the second-order expansion for $c\left(y;\theta\right)$.

Clearly,

$$\frac{d\left[\int_\theta \hat{T}\left(\hat{y}\left(\theta; y^*, \epsilon, \eta\right); y^*, \epsilon, \eta\right) d\mu\left(\theta\right)\right]}{d\eta}|_{\eta=0} = \frac{dA}{d\eta}|_{\eta=0} + \frac{dB}{d\eta}|_{\eta=0} + \frac{dC}{d\eta}|_{\eta=0} + \frac{dD}{d\eta}|_{\eta=0}$$

I characterize each of these terms. After doing so, one can divide by $F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)$ and take the limit as $\epsilon \to 0$ to arrive at the expression for $\lim_{\epsilon \to 0} \frac{d\hat{q}}{d\eta}|_{\eta=0}$.

Characterizing $\frac{dA}{d\eta}|_{\eta=0}$ First, I show that $\frac{dA}{d\eta}|_{\eta=0} = F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)$.

To see this, first write A by conditioning on $k\left(\theta\right)$. Formally, recall that $\mu\left(\theta\right)$ is the measure on the type space. Let $\mu_{\theta|k}\left(\theta|k\right)$ denote the measure of θ conditional on having curvature k (i.e. $c''\left(y\left(\theta\right)\right) = k$) and let $\mu_k\left(k\right)$ denote the measure of those having curvature k.[72] Then,

$$A = -\eta \int_k \int_{\theta|k} 1\left\{y\left(\theta\right) \in \left(y^* - \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k}}, y^* + \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k}}\right)\right\} d\mu_{\theta|k}\left(\theta|k\left(\theta\right) = k\right) d\mu_k\left(k\right)$$

Taking a derivative yields

$$\frac{dA}{d\eta} = -F\left(y^* + \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k}}\right) + F\left(y^* - \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k}}\right) - \int_k \eta\left[f_{y|k}\left(y^* + \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k}}|k\right)\sqrt{\frac{1}{2\eta k}} - f_{y|k}\left(y^* - \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k}}|k\right)\sqrt{\frac{1}{2\eta k}}\right] d\mu_k\left(k\right)$$

where $f_{y|k}\left(y|k\right)$ is the density of $y\left(\theta\right)$ given $k\left(\theta\right)$. Note that one can re-write the second term in a manner that makes it clear that it is proportional to $\sqrt{\eta}$:

$$\frac{dA}{d\eta} = -F\left(y^* + \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k}}\right) + F\left(y^* - \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k}}\right) - \sqrt{\eta}\left[\int_k \left[f_{y|k}\left(y^* + \frac{\epsilon}{2} + \sqrt{\frac{2\eta}{k}}|k\right)\sqrt{\frac{1}{2k}} - f_{y|k}\left(y^* - \frac{\epsilon}{2} - \sqrt{\frac{2\eta}{k}}|k\right)\sqrt{\frac{1}{2k}}\right] d\mu_k\left(k\right)\right]$$

Therefore, evaluating at $\eta = 0$ yields

$$\frac{dA}{d\eta}|_{\eta=0} = -\left[F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)\right]$$

[72] In other words, for any function of the type space and level of curvature, $r\left(\theta, k\left(\theta\right)\right)$, one has

$$\int\int r\left(\theta, k\left(\theta\right)\right) d\mu_{\theta|k}\left(\theta|k\left(\theta\right)\right) d\mu_k\left(k\left(\theta\right)\right) = \int r\left(\theta, k\left(\theta\right)\right) d\mu\left(\theta\right)$$

so that one can either integrate over θ (RHS) or one can first condition on curvature (and integrate over θ given curvature $k\left(\theta\right)$) and then integrate over curvature, $k\left(\theta\right)$.

Characterizing $\frac{dB}{d\eta}|_{\eta=0}$ To see this, note that

$$\begin{aligned}\frac{dB}{d\eta} &= \frac{d}{d\eta}\int_k\int_{\theta|k}\left(T\left(y^*+\frac{\epsilon}{2}\right)-T\left(y\left(\theta\right)\right)\right)1\left\{y\left(\theta\right)\in\left(y^*+\frac{\epsilon}{2},y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}\right)\right\}d\mu_{\theta|k}\left(\theta|k\left(\theta\right)=k\right)d\mu_k\left(k\right)\\ &= \int_k\left(T\left(y^*+\frac{\epsilon}{2}\right)-T\left(y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}\right)\right)\sqrt{\frac{1}{2\eta k}}f_{y|k}\left(y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}|k\right)d\mu_k\left(k\right)\end{aligned}$$

which follows from differentiating at the upper endpoint $y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}$ after conditioning on curvature k. Re-writing yields

$$\frac{dB}{d\eta} = \int_k\frac{T\left(y^*+\frac{\epsilon}{2}\right)-T\left(y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}\right)}{\sqrt{\frac{2\eta}{k}}}\frac{1}{k}f_{y|k}\left(y^*+\frac{\epsilon}{2}+\sqrt{\frac{2\eta}{k}}|k\right)d\mu_k\left(k\right)$$

Now, evaluating as $\eta\to 0$, yields

$$\begin{aligned}\frac{dB}{d\eta}|_{\eta=0} &= -T'\left(y^*+\frac{\epsilon}{2}\right)\int_k\frac{f_{y|k}\left(y^*+\frac{\epsilon}{2}|k\right)}{k}d\mu_k\left(k\right)\\ &= -T'\left(y^*+\frac{\epsilon}{2}\right)E\left[\frac{1}{k\left(\theta\right)}|y\left(\theta\right)=y^*+\frac{\epsilon}{2}\right]f\left(y^*+\frac{\epsilon}{2}\right)\end{aligned}$$

so that tax revenue is decreased by individuals decreasing their income down to $y^*+\frac{\epsilon}{2}$ in order to get the η transfer.

Characterizing $\frac{dC}{d\eta}|_{\eta=0}$ Analogous to the calculation for $\frac{dB}{d\eta}|_{\eta=0}$, it is possible to show that

$$\frac{dC}{d\eta}|_{\eta=0} = T'\left(y^*-\frac{\epsilon}{2}\right)E\left[\frac{1}{k\left(\theta\right)}|y\left(\theta\right)=y^*-\frac{\epsilon}{2}\right]f\left(y^*-\frac{\epsilon}{2}\right)$$

so that tax revenue is increased because individuals move from below $y^*-\frac{\epsilon}{2}$ up to $y^*-\frac{\epsilon}{2}$ in order to get the η transfer.

Characterizing $\frac{dD}{d\eta}|_{\eta=0}$ Finally, I show that

$$\frac{dD}{d\eta}|_{\eta=0} = \left[F\left(y^*+\frac{\epsilon}{2}\right)-F\left(y^*-\frac{\epsilon}{2}\right)\right]E\left[\frac{dy}{d\eta}T'\left(y\left(\theta\right)\right)|y\left(\theta\right)\in\left[y^*-\frac{\epsilon}{2},y^*+\frac{\epsilon}{2}\right]\right]$$

so that $\frac{dD}{d\eta}|_{\eta=0}$ is proportional to the average income effects near y^*.

To see this, note that

$$\frac{dD}{d\eta} = \frac{d}{d\eta}\int\left[T\left(\hat{y}\right)-T\left(y\right)\right]1\left\{y\left(\theta\right)\in\left(y^*-\frac{\epsilon}{2},y^*+\frac{\epsilon}{2}\right)\right\}dF\left(\theta\right)$$

Note that for these individuals in the ϵ region near y^* they only receive an income effect from the policy change. Therefore, we have

$$\frac{dD}{d\eta}|_{\eta=0} = \int T'(y(\theta)) \frac{d\hat{y}}{d\eta}|_{\eta=0} 1\left\{y(\theta) \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)\right\} dF(\theta)$$

where $\frac{d\hat{y}}{d\eta}|_{\eta=0}$ is the effect of an additional dollar of after-tax income on labor supply. One can define the income elasticity by multiplying by the after-tax price,

$$\zeta(\theta) = \left(1 - T'(y)\right) \frac{d\hat{y}}{d\eta}|_{\eta=0}$$

so that

$$\frac{dD}{d\eta}|_{\eta=0} = \int \frac{T'(y(\theta))}{1 - T'(y(\theta))} \zeta(\theta) 1\left\{y(\theta) \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right)\right\} dF(\theta)$$

Taking $\epsilon \to 0$ Now, to take the limit as $\epsilon \to 0$, note that

$$\frac{dB}{d\eta}|_{\eta=0} + \frac{dC}{d\eta}|_{\eta=0} = -T'\left(y^* + \frac{\epsilon}{2}\right) E\left[\frac{1}{k(\theta)}|y(\theta) = y^* + \frac{\epsilon}{2}\right] f\left(y^* + \frac{\epsilon}{2}\right) + T'\left(y^* - \frac{\epsilon}{2}\right) E\left[\frac{1}{k(\theta)}|y(\theta) = y^* - \frac{\epsilon}{2}\right] f\left(y^* - \frac{\epsilon}{2}\right)$$

so that

$$\lim_{\epsilon\to 0} \frac{\frac{dB}{d\eta}|_{\eta=0} + \frac{dC}{d\eta}|_{\eta=0}}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} =$$

$$= \lim_{\epsilon\to 0} \left(\frac{\epsilon}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)}\right) \left(\frac{-T'\left(y^* + \frac{\epsilon}{2}\right) E\left[\frac{1}{k(\theta)}|y(\theta) = y^* + \frac{\epsilon}{2}\right] f\left(y^* + \frac{\epsilon}{2}\right) + T'\left(y^* - \frac{\epsilon}{2}\right) E\left[\frac{1}{k(\theta)}|y(\theta) = y^* - \frac{\epsilon}{2}\right] f\left(y^* - \frac{\epsilon}{2}\right)}{\epsilon}\right)$$

or

$$\lim_{\epsilon\to 0} \frac{\frac{dB}{d\eta}|_{\eta=0} + \frac{dC}{d\eta}|_{\eta=0}}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} = \frac{1}{f(y^*)} \left(-\frac{d}{dy}|_{y=y^*} \left[T'(y) E\left[\frac{1}{k(\theta)}|y(\theta) = y\right] f(y)\right]\right)$$

Now, note also that

$$\lim_{\epsilon\to 0} \frac{-\frac{dA}{d\eta}|_{\eta=0}}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} = 1$$

and

$$\lim_{\epsilon\to 0} \frac{-\frac{dD}{d\eta}|_{\eta=0}}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} = -T'(y^*) E\left[\frac{d\hat{y}}{d\eta}|_{\eta=0}|y(\theta) = y^*\right]$$

which is given by the average income effect at y^* multiplied by the marginal tax rate.

Combining,

$$\lim_{\epsilon\to 0} \frac{d\hat{q}(y^*, \epsilon, \eta)}{d\eta}|_{\eta=0} = 1 + \frac{1}{f(y^*)} \frac{d}{dy}|_{y=y^*} \left[T'(y) E\left[\frac{1}{k(\theta)}|y(\theta) = y\right] f(y)\right] - \frac{T'(y)}{1 - T'(y)} E\left[\zeta(\theta)|y(\theta) = y\right]$$

Replacing curvature with compensated elasticity Now, note that the curvature, k, is related to the compensated elasticity of earnings. To see this, note that

$$c'(y(\theta); \theta) = 1 - \tau$$

where τ is the marginal tax rate faced by the individual, $\tau = T'(y(\theta))$. Totally differentiating with respect to one minus the marginal tax rate yields

$$c''(y(\theta))\frac{dy^c}{d(1-\tau)} = 1$$

where $\frac{dy^c}{d(1-\tau)}$ is the compensated response to an increase in the marginal keep rate, $1-\tau$. Re-writing,

$$\frac{dy^c}{d(1-\tau)} = \frac{1}{c''(y(\theta))}$$

Intuitively, the size of a compensated response to a price change is equal to the inverse of the curvature of the indifference curve.

Now, by definition, the compensated elasticity of earnings is given by

$$\epsilon^c(\theta) = \frac{dy^c}{d(1-\tau)}\frac{(1-\tau)}{y(\theta)} = \frac{1}{c''(y)}\frac{1-\tau}{y}$$

or

$$\frac{1}{k(\theta)} = \epsilon^c(\theta)\frac{y(\theta)}{1-T'(y(\theta))}$$

where $\epsilon^c(\theta)$ is the compensated elasticity of type θ defined locally around the status quo tax schedule.

Replacing $\frac{1}{k(\theta)}$ in the main equation yields

$$\lim_{\epsilon\to 0}\frac{d\hat{q}(y^*,\epsilon,\eta)}{d\eta}|_{\eta=0} = 1 + \frac{1}{f(y^*)}\frac{d}{dy}|_{y=y^*}\left[T'(y)E\left[\epsilon^c(\theta)\frac{y(\theta)}{1-T'(y(\theta))}|y(\theta)=y\right]f(y)\right] - \frac{T'(y)}{1-T'(y)}E[\zeta(\theta)|y(\theta)=y]$$

A.1.2 Adding a Participation Margin

Heretofore, I have ignored the potential for extensive margin responses. Put differently, I assumed everyone's intensive margin first order condition (equation (13)) held. Now, I assume that individuals may also move in and out of the labor force in response to changes in the tax schedule.

For simplicity, consider an alternative world where $y = 0$ was removed from individuals' feasibility set. Let $y^P(\theta)$ denote the earnings choice of type θ in this restricted world. Clearly, $y^P(\theta)$ solves

$$y^P(\theta) = argmax_{y>0}\, u(y - T(y), y; \theta)$$

For all types in the labor force in the status quo world, $y^P(\theta) = y(\theta)$. For those out of the labor force, $y(\theta) = 0$. I retain the assumption that preferences are convex over the region $y > 0$. Therefore, $y^P(\theta)$ is continuously differentiable in response to changes in the tax schedule, T. So, I allow for discrete moves between 0 and $y > 0$, but do not allow discrete moves across two different labor supply points in response to small changes in the tax schedule.

Given $y^P(\theta)$, let $c^P(\theta)$ denote the consumption level required by type θ to enter into the labor force to earn $y^P(\theta)$:

$$u\left(c^P(\theta), y^P(\theta); \theta\right) = u(y - T(0), 0; \theta)$$

Given $y^P(\theta)$ and $c^P(\theta)$, one can define the labor force participation rate at each point along the income distribution. Note that an individual of type θ chooses to work whenever

$$c^P(\theta) \leq y^P(\theta) - T\left(y^P(\theta)\right)$$

For any consumption and income level, (c, y), let $LFP(c, y)$ denote the fraction of individuals with $y^P(\theta) = y$ who choose to work, $y(\theta) = y$:

$$LFP(c, y) = \int 1\left\{c \geq c^P(\theta)\right\} d\mu\left(\theta | y^P(\theta) = y\right)$$

With this definition, one can write

$$\hat{q}(y^*, \epsilon, \eta) = A + B + C + D + P + o(\eta)$$

where P is the cost resulting from non-marginal changes in labor supply and $\frac{dP}{d\eta}|_{\eta=0,\epsilon=0}$ is given by

$$\begin{aligned}
\frac{dP}{d\eta}|_{\eta=0} &= \frac{d}{d\eta}|_{\eta=0} \int_{y^P(\theta)\in\left[y^*-\frac{\epsilon}{2}, y^*+\frac{\epsilon}{2}\right]} \left[\left(T\left(y^P(\theta)\right) - \eta - T(0)\right) LFP\left\{y^P - T\left(y^P\right) + \eta, y^P(\theta)\right\} dF(\theta)\right] \\
&= \int_{y(\theta)\in\left[y^*-\frac{\epsilon}{2}, y^*+\frac{\epsilon}{2}\right]} \left[\left(T(y(\theta)) - T(0)\right) \frac{dLFP\left\{y(\theta) - T(y(\theta)), y(\theta)\right\}}{dc} dF(\theta)\right]
\end{aligned}$$

so that

$$\begin{aligned}
\lim_{\epsilon\to 0} \frac{1}{\epsilon}\frac{dP}{d\eta}|_{\eta=0} &= E\left[\left(T(y) - T(0)\right) \frac{dLFP(y)}{dc} | y^P(\theta) = y\right] LFP(y) \\
&= \frac{T(y) - T(0)}{y - T(y)} \hat{\epsilon}\left(y^P\right) \\
&= \frac{T(y) - T(0)}{y - T(y)} \epsilon_c^{LFP}(y)
\end{aligned}$$

where $\hat{\epsilon}_c^{LFP}(y)$ is the semi-elasticity of labor force participation at y off of the base of all potential people who have $y^P(\theta)$ as their most preferred earnings point. To align with A-D, we need to replace the distribution of y^P with the distribution of y, so that we must divide by LFP. Dividing by $LFP(y)$, this is equal to the elasticity of labor force participation at $y^P(\theta)$

$$\epsilon_c^{LFP}(y) = \frac{1}{LFP(y - T(y), y)} \frac{\partial LFP(y - T(y), y)}{\partial c}$$

Therefore, we have

$$\lim_{\epsilon\to 0} \frac{d\hat{q}(y^*, \epsilon, \eta)}{d\eta}|_{\eta=0} = 1 + \frac{1}{f(y^*)} \frac{d}{dy}|_{y=y^*} \left[\frac{T'(y)}{1 - T'(y)} \epsilon^c(y) y f(y)\right] - \frac{T'(y)}{1 - T'(y)} \zeta(y) + \frac{T(y) - T(0)}{y - T(y)} \epsilon_c^{LFP}(y) \tag{14}$$

where

$$\epsilon^{c}(y) = E\left[\epsilon^{c}(\theta)\,|y(\theta) = y\right]$$

and

$$\zeta(y) = E\left[\zeta(\theta)\,|y(\theta) = y\right]$$

Multiple tax schedules Finally, one can extend these marginal costs by integrating over tax schedules, j. To see this, note that equation (14) captures the marginal cost for a change to a given tax schedule. Therefore, one can simply average across each tax schedule to obtain the average cost across tax schedules. Importantly, because j is taken to be exogenous, one does not need to account for movement across tax schedules in calculating this cost.

A.2 Proof of Proposition 2

Statement of Proposition *If $S^{ID} < 0$, there exists an $\tilde{\epsilon} > 0$ such that for any $\epsilon < \tilde{\epsilon}$ there exists an augmentation to the tax schedule in the status quo environment that generates surplus, $s_{\epsilon}^{t}(\theta)$, that is higher at all points of the income distribution: $E\left[s_{\epsilon}^{t}(\theta)\,|y(\theta) = y\right] > E\left[s_{\epsilon}(\theta)\,|y(\theta) = y\right]$ for all y. Conversely, if $S^{ID} > 0$, no such $\tilde{\epsilon}$ exists.*

Proof Suppose $S^{ID} < 0$. Then,

$$\int s(\theta)\, g(y(\theta))\, d\mu(\theta) < 0$$

so that

$$\int s_{\epsilon}(\theta)\, g(y(\theta))\, d\mu(\theta) = \epsilon \int s(\theta)\, g(y(\theta))\, d\mu(\theta) < 0$$

For any tax schedule $\hat{T}$, let $y\left(\theta;\hat{T}\right)$ denote the choice of earnings by type θ facing tax schedule $\hat{T}$. Given these choices, total tax revenue is given by

$$R\left(\hat{T}\right) = \int \hat{T}\left(y\left(\theta;\hat{T}\right)\right) d\mu(\theta)$$

Now, consider an augmented tax schedule. Let $P = \{P_j\}_{j=1}^{N_P}$ denote a partition of the income distribution into intervals and let η_j^P denote transfers provided to each such region of the income distribution:

$$\hat{T}_{\epsilon}^{P}(y) = T(y) - \epsilon \sum_{j=1}^{N_P} \eta_j^P 1\{y \in P_j\}$$

and let

$$\hat{T}_{\epsilon}^{j}(y) = T(y) - \epsilon \eta_j^P 1\{y \in P_j\}$$

I assume that

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{P}\right)=\sum_{j=1}^{N_P}\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{j}\right) \tag{15}$$

This assumption is satisfied for most common forms of preferences. It would be violated if, for example, there were a mass of agents perfectly indifferent between three earnings points. Then, providing additional transfers to one of these two points would both induce movement from the other point and thus the sum of the two tax movements would be larger than the combined tax movement. In practice, it is satisfied if agents have at most two points of indifference in their utility function (since then their movement is affected by the relative price changes to be in these two locations).

Note that each partition can be represented as

$$P_j=[y_j^*-\epsilon_j,y_j^*+\epsilon_j)$$

so that

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{j}\right)=-\eta_j^P\frac{d}{d\eta}|_{\eta=0}\hat{q}\left(y_j^*,\epsilon_j,\eta\right)\Pr\left\{y\left(\theta\right)\in P_j\right\}$$

where $\Pr\left\{y\left(\theta\right)\in P_j\right\}=\mu\left\{y^{-1}\left(P_j\right)\right\}$.

Now, define η_j^P as

$$\eta_j^P=\sup\left\{E\left[s\left(\theta\right)|y\left(\theta\right)=y\right]|y\in P_j\right\}-\frac{S^{ID}}{2}\bar{h}$$

where $\bar{h}=E\left[h\left(\theta\right)\right]$ is the average value of the marginal cost of taxation. Let $s_{\epsilon}^{t}\left(\theta\right)$ denote the surplus the individual obtains when facing tax schedule $\hat{T}_{\epsilon}^{j}$. By the envelope theorem (and the assumption of no externalities / GE effects), there exists $\tilde{\epsilon}$ such that for all $\epsilon<\tilde{\epsilon}$, the individual obtains surplus at least as large as $\epsilon\left(\sup E\left[s\left(\theta\right)|y\left(\theta\right)=y,y\in P_j\right]\right)$

$$E\left[s_{\epsilon}^{t}\left(\theta\right)|y\left(\theta\right)=y\right]>\epsilon\left(\sup E\left[s\left(\theta\right)|y\left(\theta\right)=y|y\in P_j\right]\right) \tag{16}$$

for all $\epsilon\in(0,\tilde{\epsilon})$ (to see this, note that the tax augmentation not only gives people surplus s^y but also provides $-\frac{S^{ID}}{2}>0$; so this inequality is made strict).

Now, consider the marginal cost of the policy. By construction

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{P}\right)=-\sum_{j=1}^{N_P}\eta_j^P\frac{d}{d\eta}|_{\eta=0}\hat{q}\left(y_j^*,\epsilon_j,\eta\right)\Pr\left\{y\left(\theta\right)\in P_j\right\}$$

and taking the limit as partition widths go to zero,

$$\lim_{width(P)\to 0}\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{P}\right)=-\lim_{width(P)\to 0}\sum_{j=1}^{N_P}\left(\sup\left\{E\left[s\left(\theta\right)|y\left(\theta\right)=y\right]|y\in P_j\right\}-\frac{S^{ID}}{2}\right)\frac{d}{d\eta}|_{\eta=0}\hat{q}\left(y_j^*,\epsilon_j,\eta\right)\Pr\left\{y\left(\theta\right)\in P_j\right\}$$

Note that the terms inside the sum have limits that exist and are unique (because $h\left(y\right)$ is assumed to be continuous and the mean surplus function is assumed to be continuous). Note in principle this limit existing does not require continuity of either the surplus function or the marginal cost function

$h(y)$ – some suitable integrability condition would work – but this is sufficient. So,

$$\begin{aligned}
\lim_{width(P)\to 0}\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_\epsilon^P\right) &= -\int s^Y\left(y(\theta)\right)h\left(y(\theta)\right)d\mu(\theta)+\frac{S^{ID}}{2}\bar{h}\\
&= -S^{ID}\bar{h}+\frac{S^{ID}}{2}\bar{h}\\
&= -\frac{S^{ID}}{2}\bar{h}
\end{aligned}$$

which is positive. Therefore, there exists $\epsilon^* < \tilde{\epsilon}$ such that for all $\epsilon \in (0,\epsilon^*)$ we have $R\left(\hat{T}_\epsilon^P\right) > 0$ and $E\left[s_\epsilon^t(\theta)\,|y(\theta)=y\right] > E\left[s(\theta)\,|y(\theta)=y\right]$ for all y.

Converse Now suppose $S^{ID} > 0$. Then,

$$\int s(\theta)\,h\left(y(\theta)\right)d\mu(\theta) > 0$$

And, suppose for contradiction that some $\tilde{\epsilon}$ exists so that there are a set of tax schedules, $\hat{T}_\epsilon$, that deliver greater surplus along the income distribution, $\frac{d}{d\epsilon}|_{\epsilon=0}E\left[s_\epsilon^t(\theta)\,|y(\theta)=y\right] \geq E\left[s(\theta)\,|y(\theta)=y\right]$. I will show that this implies the tax schedule modification is not budget neutral for sufficiently small ϵ.

Note that the envelope theorem implies $\frac{d}{d\epsilon}|_{\epsilon=0}\hat{T}_\epsilon(y) = \frac{d}{d\epsilon}|_{\epsilon=0}s_\epsilon^t(\theta)$ for all θ such that $y(\theta)=y$. For any $\epsilon>0$ and $\gamma>0$, one can approximate the revenue function using a partition $P^\gamma = \left\{P_j^\gamma\right\}_{j=1}^{N_{P^\gamma}}$ and a step tax function $T_\epsilon^{P^\gamma}$ that provides exactly $E\left[s_\epsilon^t(\theta)\,|y(\theta)\in P_j^\gamma\right]$ units of tax reduction. Therefore, the marginal cost of the policy is approximated by $\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_\epsilon^{P^\gamma}\right)$,

$$\left|\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_\epsilon\right)-\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_\epsilon^{P^\gamma}\right)\right| < \gamma$$

where

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_\epsilon^{P^\gamma}\right) = -\sum_{j=1}^{N_{P^\gamma}}\frac{d}{d\epsilon}|_{\epsilon=0}E\left[s_\epsilon^t(\theta)\,|y(\theta)\in P_j^\gamma\right]\frac{d}{d\eta}|_{\eta=0}\hat{q}\left(y_j^*,\epsilon_j,\eta\right)\Pr\left\{y(\theta)\in P_j\right\}$$

where $P_j^\gamma = [y_{\gamma,j}^*-\epsilon_{\gamma,j}, y_{\gamma,j}^*+\epsilon_{\gamma,j})$. For sufficiently small ϵ we know that $E\left[s_\epsilon^t(\theta)\,|y(\theta)\in P_j^\gamma\right] > E\left[s_\epsilon(\theta)\,|y(\theta)\in P_j^\gamma\right]$. Therefore,

$$\frac{d}{d\epsilon}|_{\epsilon=0}E\left[s_\epsilon^t(\theta)\,|y(\theta)\in P_j^\gamma\right] > \frac{d}{d\epsilon}|_{\epsilon=0}E\left[s_\epsilon(\theta)\,|y(\theta)\in P_j^\gamma\right]-\gamma$$

since both have values of zero when $\epsilon=0$.

Therefore,

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}^{P}\right) < -\sum_{j=1}^{N}\frac{d}{d\epsilon}|_{\epsilon=0}E\left[s_{\epsilon}(\theta)\,|y(\theta)\in P_j^{\gamma}\right]\frac{d}{d\eta}|_{\eta=0}\hat{q}\left(y_j^*,\epsilon_j,\eta\right)\Pr\left\{y(\theta)\in P_j^{\gamma}\right\}+\gamma$$

and taking the limit as the partition widths converge towards zero (so that $\gamma\to 0$), we arrive as

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}\right)\leq -S^{ID}E\left[h(\theta)\right]<0$$

so that the policy is not budget neutral.

Discussion The proof relied on two key assumptions. First, I assume that providing a small amount of money through modifications in the tax schedule generates surplus of at least the mechanical amount of money provided in the absence of any behavioral response. This follows from the envelope theorem, combined with the assumption that infinitesimal tax changes in one portion of the income distribution do not affect the welfare of anyone at other points of the distribution. This was implicitly assumed by writing the utility function as a function of one's own consumption and earnings, and not a function of anyone else's choices of labor supply or earnings. For example, if taxing the rich caused them to reduce their earnings which in turn increased the wages of the poor, then equation (16) would no longer hold, since individuals outside of the intended target of the tax transfers would have surplus impacts. Accounting for such general equilibrium effects is an interesting and important direction for both theoretical and empirical work.

Second, I assume that the revenue function is continuously differentiable and additive in modifications to the tax schedule. This is primarily a technical assumption that rules out types that are indifferent to many points along the income distribution (which would cause them to be double-counted as costs in equation (15)).

A.3 Proof of Proposition 3

Statement of Proposition Suppose Assumption 3 holds. If $S^{ID}>0$, there exists $\tilde{\epsilon}>0$ such that for any $\epsilon<\tilde{\epsilon}$, there exists an augmentation to the tax schedule in the alternative environment that delivers surplus $s_{\epsilon}^{t}(\theta)$ that is on average positive at all points along the income distribution: $E\left[s_{\epsilon}^{t}(\theta)\,|y(\theta)=y\right]>0$ *for all* y*. Conversely, if* $S^{ID}<0$*, then no such* $\tilde{\epsilon}$ *exists.*

Proof I provide the brief sketch here that does not go through the formality of defining the partitions as in the proof above, but one can do so analogously to the proof of Proposition 1. Let $y(\theta)$ continue to denote the choice of income of a type θ in the status quo environment. Let $y_{\epsilon}^{\alpha}(y)$ denote the choice of income in the alternative environment made by those who chose y in the status quo environment. Per Assumption 3, this function is a bijection. Given the surplus function, $E\left[s_{\epsilon}(\theta)\,|y(\theta)=y\right]$, consider a modification to the income distribution that taxes away all but $\epsilon\frac{S^{ID}}{2}$ of this surplus to those earning y in the status quo (i.e. those earning $y_{\epsilon}^{\alpha}(y)$ in the ϵ-alternative environment). If $\tilde{T}_{\epsilon}$ is the tax schedule in the ϵ-alternative environment, then the modified tax schedule is

$$\hat{T}_\epsilon(y) = \tilde{T}_\epsilon(y) + \epsilon\left(E\left[s(\theta)\,|y(\theta) = (y^\alpha_\epsilon)^{-1}(y)\right] - \frac{S^{ID}}{2}\right)$$

Let $s^t_\epsilon(\theta)$ denote the surplus of the tax-modified ϵ-alternative environment with tax schedule $\hat{T}_\epsilon(y)$. For sufficiently small ϵ, the off-setting transfer ensures everyone is better off relative to the status quo (note this relies on the fact that $S^{ID} > 0$, so that there is aggregate inequality deflated surplus to spread around). Hence, $s^t_\epsilon(\theta) > s_\epsilon(\theta) - E[s_\epsilon(\theta)\,|y(\theta) = y]$ for sufficiently small ϵ; and taking the expectation conditional on $y(\theta) = y$ yields

$$E\left[s^t_\epsilon(\theta)\,|y(\theta) = y\right] > 0 \quad \forall y$$

Now, one needs to show that, for sufficiently small ϵ, the cost of the modification to the tax schedule is not budget-negative. Note that for each y, the tax modification provides a transfer of $E\left[s_\epsilon(\theta)\,|y(\theta) = (y^\alpha_\epsilon)^{-1}(y)\right]$. Note that Assumption 3, the marginal cost of implementing these surplus transfers is the same as in the status quo environment.

$$\frac{dR}{d\epsilon}|_{\epsilon=0} = \int E[s(\theta)\,|y(\theta) = y]\,g(y)\,dF(y) - \frac{S^{ID}}{2} = \frac{S^{ID}}{2} > 0$$

so that the transfer scheme is feasible for sufficiently small ϵ.

A.4 Proof of Proposition 4

Statement of Proposition Suppose $\overline{S}^{ID} < 0$. Then, there exists an $\tilde{\epsilon} > 0$ such that, for each $\epsilon < \tilde{\epsilon}$ there exists a modification to the income tax schedule that delivers a Pareto improvement relative to $s_\epsilon(\theta)$. Conversely, if $\overline{S}^{ID} > 0$, there exists an $\tilde{\epsilon} > 0$ such that for each $\epsilon < \tilde{\epsilon}$ any budget-neutral modification to the tax schedule results in lower surplus for some θ relative to $s_\epsilon(\theta)$.

Proof The proof follows immediately by providing surplus $\overline{s}_\epsilon(y) = \sup\{s_\epsilon(\theta)\,|y(\theta) = y\}$ instead of $E[s_\epsilon(\theta)\,|y(\theta) = y]$ in the proof of Proposition 2.

A.5 Proof of Proposition 5

Statement of Proposition Suppose 3 holds. Suppose $\underline{S}^{ID} > 0$. Then, there exists an $\tilde{\epsilon} > 0$ such that, for each $\epsilon < \tilde{\epsilon}$ there exists a modification to the income tax schedule in the alternative environment that delivers positive surplus to all types, $s^t_\epsilon(\theta) > 0$ for all θ.

Proof The proof follows immediately by providing surplus $\underline{s}_\epsilon(y) = \inf\{s_\epsilon(\theta)\,|y(\theta) = y\}$ instead of $E[s_\epsilon(\theta)\,|y(\theta) = y]$ in the proof of Proposition 3.

A.6 Distribution and Surplus

Here, I show that $r(\alpha) = \alpha$ minimizes the surplus of the alternative environment when making distributional comparisons when $g(\alpha)$ is decreasing. More formally, I show that $r(\alpha) = \alpha$ (a.e.) minimizes the surplus of the alternative environment:

$$\int_0^1 \left(Q_a\left(r\left(\alpha\right)\right) - Q_0\left(\alpha\right)\right) g\left(\alpha\right) d\alpha \leq \int_0^1 \left(Q_a\left(\alpha\right) - Q_0\left(\alpha\right)\right) g\left(\alpha\right) d\alpha$$

The proof goes as follows. First, start with an r such that $r(\alpha) \neq \alpha$ a.e. One can construct $\hat{r}$ that delivers lower inequality deflated surplus by undoing some of the crossed paths and replacing a portion of $r(\alpha)$ with $r(\alpha) = \alpha$. Heuristically, this lowers surplus to individuals at the bottom and raises surplus at the top. Because g is decreasing, the increase in surplus at the top is valued less than the decrease in surplus at the bottom. So, $\hat{r}$ will have lower inequality deflated surplus.

Formally, there exists $\epsilon > 0$ and α_1 such that

$$\alpha_1 = \inf\left\{int\left\{\alpha | r\left(\alpha\right) \neq \alpha\right\}\right\}$$

and

$$\left(\alpha_1, \alpha_1 + \epsilon\right) \subset int\left\{\alpha | r\left(\alpha\right) \neq \alpha\right\}$$

Intuitively, α_1 is the minimum quantile at which $r(\alpha) \neq \alpha$ for a non-trivial set of points. Now, define $\hat{r}(\alpha)$ to be the permutation that instead allows $r(\alpha) = \alpha$ on the interval $(\alpha_1, \alpha_1 + \epsilon)$ but then fills in the points $\{r(\alpha) | \alpha \in (\alpha_1, \alpha_1 + \epsilon)\}$ using the set $\{\alpha | r(\alpha) \in (\alpha_1, \alpha_1 + \epsilon)\}$. Then, the difference between r and $\hat{r}$ is a permutation cycle between $(\alpha_1, \alpha_1 + \epsilon)$ and higher values of α. This means the experiment r offers less surplus to types at the bottom of the distribution but the same total surplus. Because g is decreasing, this implies r has lower surplus than $\hat{r}$.

B Summary Statistics and Estimation Details

Appendix Table I presents the summary statistics of the sample used to construct the estimates of the shape of the income distribution conditional on the marginal income tax rate. Overall, there are roughly 100M filers aged 25-60 used in the analysis, with mean family incomes of roughly \$65M, and mean ordinary incomes of \$46M. The negative tax rates are generated by the various phase-in regions of the EITC for filers of different marital status and combinations of children and EITC-eligible children.[73]

[73] As is well-known, EITC dependents is technically different from the number of dependents one can claim for baseline 1040 deductions. Fortunately, the tax data contains information on both the number of EITC-eligible dependents and the number of 1040 dependents, allowing me to precisely identify the marginal tax rate.

Appendix Table I
Summary Statistics

Marginal Federal Tax Rate (1)	Number of Filers (2)	Percent of Filers (3)	Mean Ordinary Income (4)	Mean Family Income (5)
-35.0%	589,750	0.6%	79	11,622
-30.0%	1,413,594	1.4%	229	11,598
-25.0%	6,604	0.0%	16,004	34,459
-24.0%	1,366,424	1.4%	346	9,435
-19.0%	10,905	0.0%	15,088	30,532
2.4%	2,738,247	2.7%	170	6,320
7.4%	18,321	0.0%	19,049	35,938
10.0%	19,355,436	19.3%	(7,140)	18,754
15.0%	35,974,064	35.9%	32,510	52,689
17.4%	620	0.0%	58,739	65,642
17.7%	2,736,669	2.7%	1,120	11,585
22.7%	927	0.0%	12,936	16,493
25.0%	19,217,503	19.2%	74,974	97,153
26.0%	3,035,148	3.0%	6,416	25,402
28.0%	6,266,531	6.3%	193,243	234,677
31.0%	1,078,152	1.1%	15,822	32,212
31.1%	4,868,642	4.9%	4,622	28,992
33.0%	225,353	0.2%	243,716	234,349
35.0%	400,306	0.4%	1,977,424	1,154,254
36.1%	795,541	0.8%	16,523	38,355
Other	549	0.0%	15,934,759	54,534
Total	100,099,286	100.0%	46,110	64,745

Notes. This table presents summary statistics for the universe of income tax returns in 2012 for U.S. citizens aged 25-60. This sample is used to construct the elasticity of the density of the income distribution (i.e. alpha) for each marginal tax rate. The table presents the number of filers, mean ordinary income, and mean family income by each federal marginal tax rate. The mean federal tax rate is the effective marginal tax rate each filer would face on an additional dollar of income. This equals the tax on ordinary income for most filers, but includes additional tax rates generated by the earned income tax credit (EITC) for EITC filers and the alternative minimum tax (AMT) for filers subject to the AMT.

To estimate the Pareto parameter of the income distribution, I proceed as follows. First, for computational simplicity, I define 1000 equally sized bins of ordinary income. I then collapse the data to generate counts of returns in each of these 1000 bins separately for returns facing different tax schedules, j. I generate these groups as the intersection of filing status, EITC status (marital status + number of qualified EITC dependents), and those subject to the alternative minimum tax rate.

Given these groupings, I estimate α_j in a manner that allows it to vary with the marginal tax rate for a majority of the population. For tax schedules with at least at least 500,000 observations with earnings between the 10th and 99th percentile of the income distribution, I estimate the elasticity of the income distribution separately for each filing characteristic.[74] To do so, I construct the log density of the income distribution measuring the number of households in each bin divided by the width of the bin. I then regress this on a fifth order polynomial of log income in the bin (where income is the mean income within the bin). The estimated slope at each bin generates an estimate of α_j for each income bin in tax group j. I verify that the results are virtually identical when increasing or decreasing the number of bins or changing the number of polynomials in the regression.

For the remaining smaller tax groups (~25% of the sample) with fewer than 500,000 returns, I impose the assumption that the elasticity of the income distribution is the same across these less-populated tax schedules at a given level of income.[75] I then take advantage of the fact that the aggregate elasticity can be written as a weighted average of the elasticities of the income distribution for each marginal tax rate, α_j. So, I estimate the elasticity of the aggregate income distribution and then construct the implied elasticity for these smaller groups as the population weighted difference between the total elasticity and the elasticities of the larger tax groups. To estimate the elasticity of the aggregate income distribution, I regress the log density on a tenth order polynomial in log income for each bin (again, results are nearly identical if one includes additional polynomials) and compute the slope at each bin.

The advantage of this estimation approach is that it allows the elasticity of the income distribution to vary non-parametrically with the tax rates for ~75% of the sample. Hence, I allow for significant correlation between the shape of the income distribution and the marginal tax rate, as is potentially required for accurate estimation of the substitution effect in equation (6) in the presence of multiple tax schedules.

For individuals near the top of the income distribution, the local calculation of the elasticity of the income distribution becomes difficult and potentially biased because of endpoint effects. Intuitively, the binning of incomes into 1,000 bins ignores the fact that the U.S. income distribution has a fairly thick upper tail. Fortunately, it is well documented that the upper tail of the income distribution is Pareto, and hence has a constant elasticity so that $\alpha(y) = \frac{E[Y|Y \geq y]-y}{y}$ (Saez (2001)). Hence, I also compute an "upper tail" value of α given by $\frac{E[Y|Y \geq y]-y}{y}$ for each income bin. Figure (16) (left panel)

[74] I do not include the returns below the 10th quantile of the income distribution because of the large fraction of returns posting exactly $3k in ordinary income, which introduces significant nonlinearities in some of these groups. Above the 99th percentile, I follow a strategy from Saez (2001) described below.

[75] This 500,000 threshold is chosen for computational simplicity on the remaining groups, but the results are similar to lowering it to 250,000.

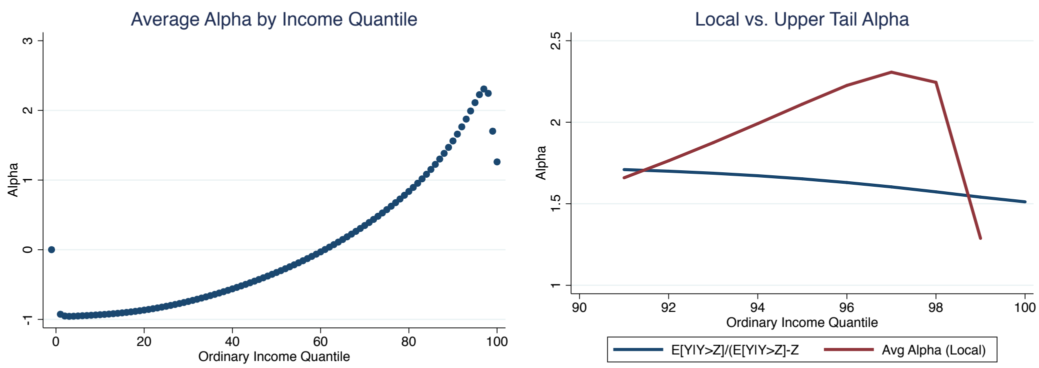

Figure 16: Alpha, $\alpha(y)$.

plots the average local estimate of α (using the fifth order polynomial) across the income distribution and Figure (16) (right panel) plots both this estimate and the upper tail value of α, $\frac{E[Y|Y\geq y]-y}{y}$, for the upper decile of the income distribution.

For the upper regions of the income distribution, the value of $\frac{E[Y|Y\geq y]-y}{y}$ converges to around 1.5, consistent with the findings of Diamond and Saez (2011) and Piketty and Saez (2012). Conversely, the local estimate of the elasticity of the income distribution arguably becomes downwardly biased in the upper region because the fifth order polynomial does not capture the size of the thick tail in the top-most income bucket. Hence, for incomes in this upper region with earnings above $250,000, I assign the maximum value of these two estimates.

C Cross Country Comparison Data

The inequality deflated surplus associated with the income distribution in country a is given by

$$S_a^{ID} = \int_0^1 \left[Q_a(\alpha) - Q_0(\alpha)\right] g^H\left(Q_0(\alpha)\right) d\alpha$$

Often, country income inequality data is available as the share of income accruing to each quantile of the income distribution (often this is used to construct a Lorentz curve). To this aim, one can express the inequality deflated surplus as

$$S_a^{ID} = Y^a \int_0^1 \frac{Q_a(\alpha)}{Y^a} g^H\left(Q_0(\alpha)\right) d\alpha - Y^0 \int_0^1 \frac{Q_0(\alpha)}{Y^0} g^H\left(Q_0(\alpha)\right) d\alpha$$

where Y^a is mean income in country a and Y^0 is mean income in the U.S., and $\frac{Q_a(\alpha)}{Y^a}$ is the income at the α quantile of the distribution relative to mean income (and can be computed from income shares of each quantile).[76] Note that if the two countries have the same distribution of income (as measured by the ratios of quantiles Q_a to mean income, Y^a) then a comparison of surplus between the two countries

[76] Because data on income distribution is generally at the household level, I use the household income deflator, g^H.

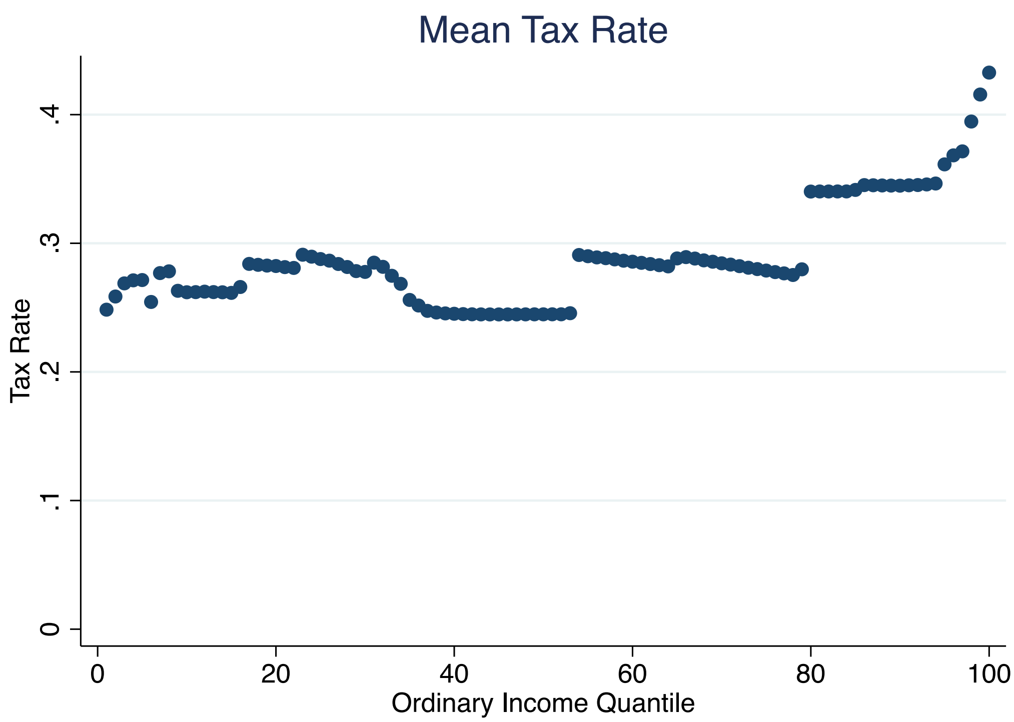

Figure 17: Mean tax rate, $\tau(y)$, by ordinary income quantile. τ is the sum of the federal income tax rate, state taxes, Medicare, sales taxes, and EITC top-up, as discussed in Section 4.1.

can be conducted using solely the difference in mean incomes. More generally, $\int_0^1 \frac{Q_a(\alpha)}{Y^a} g\left(y\left(\alpha\right)\right) d\alpha$ is a deflation factor that neutralizes distributional comparisons across countries.[77]

I estimate these deflation factors using data from the World Bank Development Indicators and UN World Income Inequality Database. These databases aggregate information from household surveys within each country and provide the share of income accruing to each quintile or decile of the household income distribution.[78] For each country, I take the most recent survey providing information on the distribution of income and limit the analysis to countries for which surveys are available post-2000. This yields a sample of 130 countries.[79]

For each country, I construct an approximation to the deflation factor using these quintiles or deciles

$$\int_0^1 \frac{Q_a\left(\alpha\right)}{Y^a} g\left(y\left(\alpha\right)\right) d\alpha \approx \sum_{q=1}^{N_q} s_q g_q$$

where $g_q = \int_{\frac{q-1}{N_q}}^{\frac{q}{N_q}} g\left(y\left(\alpha\right)\right) d\alpha$ is the average deflator for income in the qth region of the income distribution and $N_q = 5$ for countries with quintile information and $N_q = 10$ for countries with decile information. And, $s_q = \frac{1}{Y^a} E\left[Q_a\left(\alpha\right) | \frac{q-1}{N_q} \leq \alpha < \frac{q}{N_q}\right]$ is the average income of individuals in the qth region of the income distribution relative to mean income, Y_a.

In addition to income distribution data, I use data from the World Bank Development Indicators for measures of gross national income per capita in 2012 measured in U.S. dollars.[80]

Figure 18 plots the deflation factors, $\int_0^1 \frac{Q_a(\alpha)}{Y^a} g^H\left(Q_0\left(\alpha\right)\right) d\alpha$, for each country relative to their gross national income (GNI) per capita in 2012. In general, richer countries tend to have less upper tail inequality and hence have higher deflator adjustment factors. A notable exception is the U.S., whose deflation factor is below 0.9. Reflecting its larger degree of income inequality, the U.S. is an outlier in

[77] Note that both have mean 1, therefore, $\int_0^1 \frac{Q_a(\alpha)}{Y^a} g^H\left(Q_0\left(\alpha\right)\right) = 1 + cov\left(\frac{Q_a(\alpha)}{Y^a}, g\left(Q_0\left(\alpha\right)\right)\right)$. Intuitively, surplus is deflated to the extent to which it falls more heavily on those with lower values of the inequality deflator, g. Note that this covariance is zero if either (a) it is costless to redistribute across the income distribution ($g\left(y\right) = 1$) or (b) there is no income heterogeneity, $y\left(\alpha\right) = Y_a$.

[78] One downside of these surveys is that they often under-state the incomes accruing to the top of the income distribution. Although the World Top Incomes Database of Piketty and Saez (http://topincomes.g-mond.parisschoolofeconomics.eu) contains information on top incomes for the top deciles, they do not provide the top quintile incomes, which would be required to adequately merge these with the income distribution estimates by quintile. Hence, I do not perform adjustments to the World Bank Development Indicators or the UN WIID database. Because the U.S. has relatively higher income shares in the top 1% relative to other countries with similar income levels, this likely under-states the degree to which the U.S.'s relative position in the world income distribution drops when deflating incomes using the inequality deflator.

[79] An important limitation of using these surveys is that they generally provide measures of the distribution of pre-tax income, whereas the most natural conceptual experiment would involve post-tax income comparisons. To the extent to which the countries that are compared to the U.S. have a more progressive tax structure, this will lead to an under-statement of the impact of the US attempting to replicate the more equal after-tax income distributions in these countries. Moreover, to the extent to which public goods are allocated more to the poor than the rich in these countries, this approach will also under-state the degree of inequality differences in a more broad measure of surplus.

[80] The World Bank adjusts for differences in price levels using the ATLAS method. In principle, one could use the inequality deflator, combined with estimates of the commodity shares across the income distribution, to formulate a new price adjustment across countries that accounts for heterogeneity in the price differences at different points of the income distribution. This is an important direction for future work. Moreover, one might prefer to use measures of household income, rather than income per capita. Using GNI per capita induces bias to the extent to which countries differ in their distribution of household size across the income distribution.

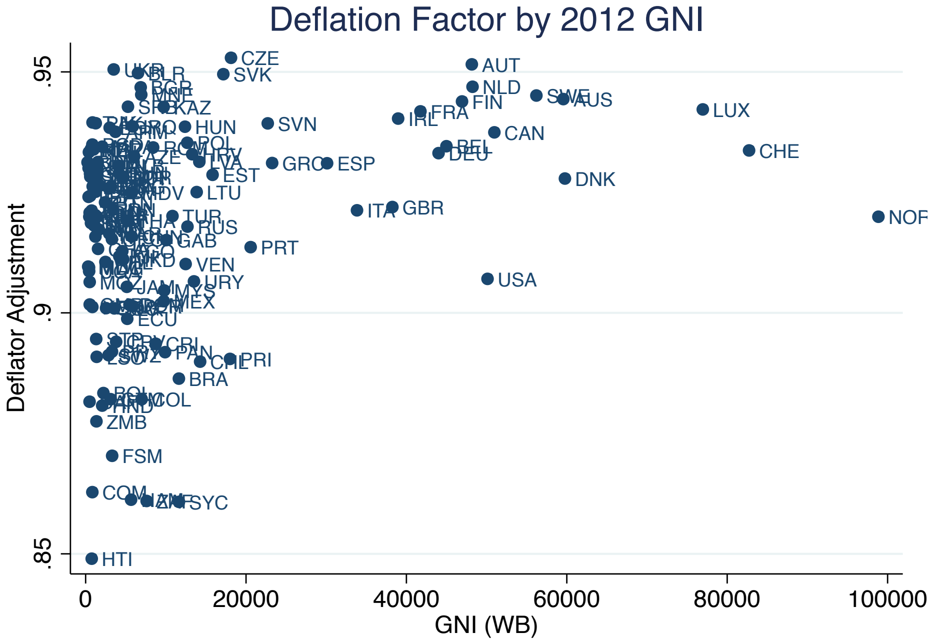

Figure 18: This figure presents the deflation factor, $\int_0^1 \frac{Q_a(\alpha)}{Y^a} g^H \left(Q_0\left(\alpha\right)\right) d\alpha$, on the vertical axis and 2012 GNI on the horizontal axis for each of the 130 countries in the sample.

the relationship between the deflation factor and income level.[81]

D Deflating changes in pre- and post-tax census income

This appendix measures the cost of changes in the pre-tax household income distribution, as opposed to the post-tax analysis in the main text. Deflating pre-tax household income corresponds to changes in welfare if one assumes that the benefits of taxation are returned to those directly paying taxes.

To construct the pre-tax series, I use the distribution of household income from the U.S. Census Bureau for incomes up to the 95th percentile. However, as is well known, survey datasets tend to under-state top incomes (Piketty and Saez (2003)). Therefore, I replace the top 5% of the income distribution using the top 5% incomes reported by Piketty and Saez (2007). I use the updated top incomes statistics through 2011. Appendix Figure 19 illustrates the difference between the top 5% measure from Piketty and Saez (2007) versus the U.S. Census Bureau. Prior to 1995, this difference is generally fairly small (e.g. $5K) but rises to upwards of $50K after 1995, consistent with the rise of extreme top incomes.

Appendix Figure 19 reports the top 5% income from the census and from Piketty and Saez. As one can see, the census tends to diverge more so over time with income differences reaching upwards

[81] To avoid bias from making comparisons between survey and administrative data, I use the income distribution series for the U.S. used by the World Inequality Database (which is the same meta-source as the data from Austria and the Netherlands).

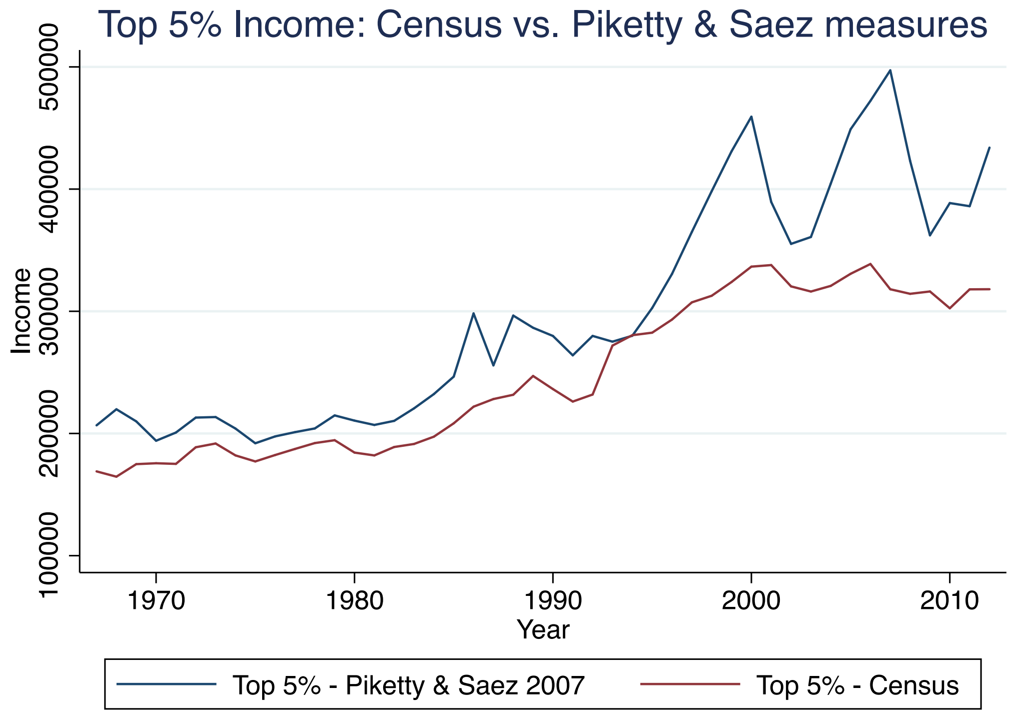

Figure 19: Top 5% Incomes by Data Source

of 50,000 after 1995.

Given the quantile points in Figure 20, I apply the deflator using the average household income deflator value in each quintile below the 80th percentile of the household income distribution and the average household income deflator between the 80th and 95th percentile and the top 5%. The deflator takes on values of roughly 1.1 for the first three quintiles, 0.97 for the fourth quantile, and 0.82 for the 80-95th percentile, and 0.66 for the top 5%.

Figure 21 plots inequality deflated surplus, S_t^{ID}, by year and compares it to the mean difference in household income between year 2011 and year t. The orange dash-dot line at the top plots the change in mean household per capita income (CPI deflated) relative to 2012. The blue solid line plots the inequality deflated value of this change for the baseline elasticity specification. The long and short dashed lines plot the low and high elasticity specification.

For the baseline elasticity specification, the social cost of increased income inequality since 1970 is roughly \$5,250, which is roughly 21.5% of the \$24K in mean increase in household income between 1970 and 2012.[82] Multiplying by 119K households, this implies a social cost of increased income inequality since 1970 of roughly \$625B. In the high elasticity specification, the social cost of inequality is nearly doubled to \$8,733, or 36% of the household income increase since 1970. In the low elasticity specification, the social cost of inequality is \$1,988, roughly 8% of the mean household income increase

[82]The combined data series that replaces the top 5% incomes with the measures of Piketty and Saez (2007) suggests household income increased from \$54,463 in 1970 to \$78,769 in 2012 (measured in 2012 dollars).

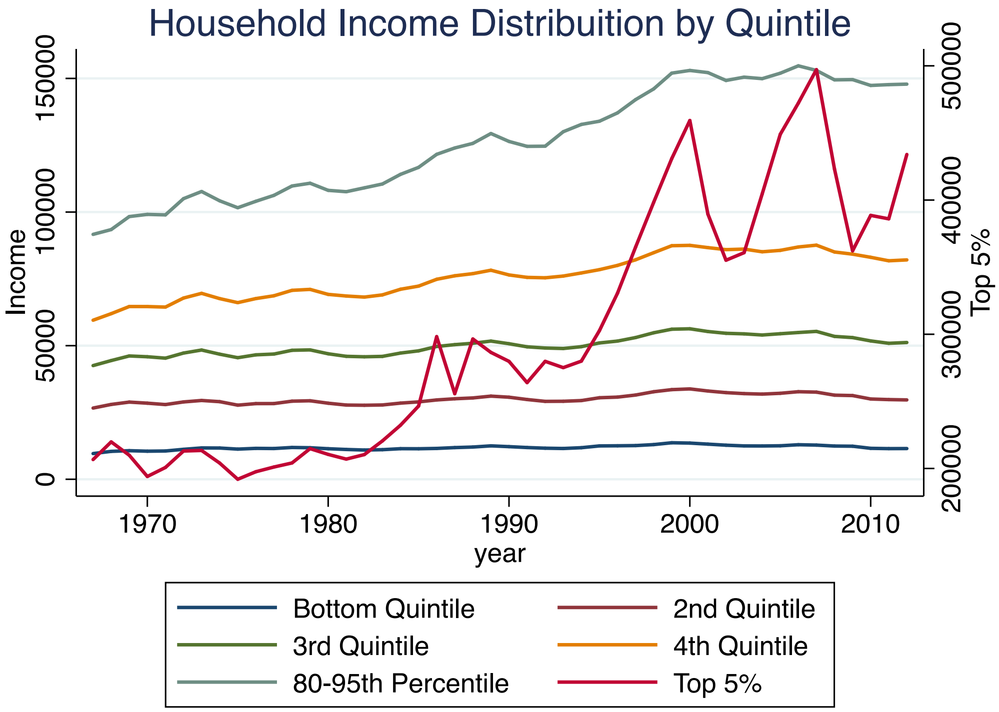

Figure 20: Income Distribution by Quintile

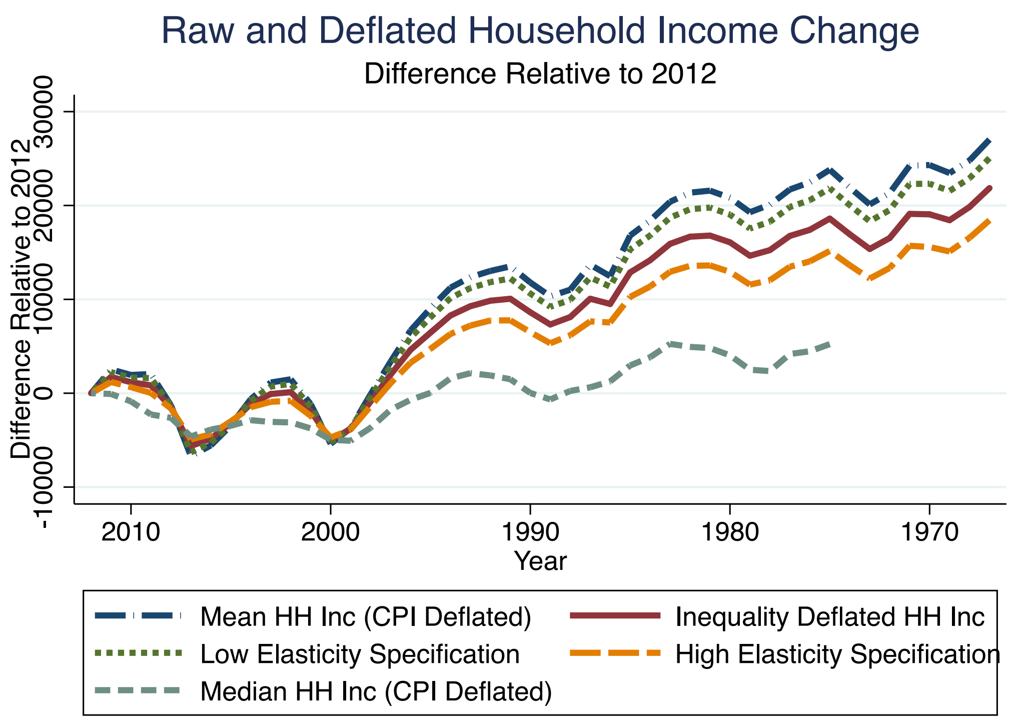

Figure 21: Raw and Deflated Household Income Change Relative to 2012

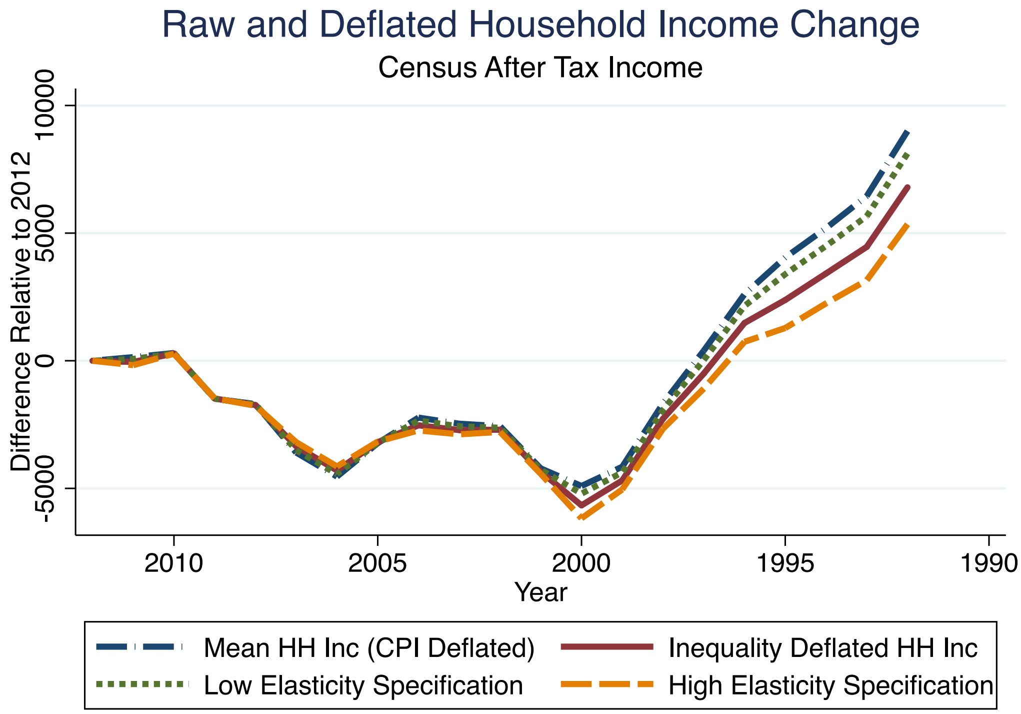

Figure 22: Raw and Deflated Household Income Change Relative to 2012

since 1970.

In addition to post-tax census data, Figure 22 plots the increase in income inequality using census micro data from 1992-2012 (using the same sample as discussed in the main text for which I construct the 2009-2012 series to extend the CBO 1979-2009 data). In general, the pattern is similar to the before tax pattern in the Census data (a cost of increased income inequality in the 1990s), but unfortunately consistent after tax measures in the CPS do not extend back before 1992. Moreover, because the CBO data relies on tax data, it arguably provides a better picture of the increased income inequality at the top if the income distribution. Therefore, I rely on the CBO data for the analysis in the main text, aside from the extension from 2009 to 2012.