Efficient Welfare Weights

Nathaniel Hendren*

August, 2019

Abstract

How should we measure economic efficiency and the desirability of economic policies? Motivated by the classic work of Kaldor (1939) and Hicks (1940), a canonical approach measures the unweighted sum of willingnesses to pay minus government costs, ignoring distributional incidence. In contrast, this paper provides efficient welfare weights that implement the Kaldor-Hicks tests for efficiency but account for the heterogenous distortionary cost of taxation. Using the universe of tax returns, I show the joint distribution of the income tax schedule and taxable earnings imply that it is efficient to weight surplus to the poor more than to the rich. Calibrating the model using elasticities of taxable income suggest surplus to the poor should be weighted 1.5-2x more than surplus to the rich. Intuitively, \$1 of surplus to the poor can be turned into \$1.5-\$2 of surplus to the rich through reductions in distortionary taxation. The weights provide an empirical welfare framework that relies on the Pareto principle and causal effects, instead of a social welfare function.

1 Introduction

Economic policies create winners and losers. In order to have a preference over policy changes, one must take a stance on how to weigh the gains to the winners against the losses to the losers.

Kaldor and Hicks classic definition of economic efficiency was originally intended to resolve this tradeoff. They proposed constructing a sum of individuals' willingnesses to pay. If this sum is negative, a set of individual-specific lump-sum transfers could reach an allocation that is *Pareto* superior to the

*Harvard University, nhendren@fas.harvard.edu. This paper is a revised version of a paper that was originally circulated in July 2014 under the title, "The Inequality Deflator: Interpersonal Comparisons without a Social Welfare Function", NBER WP No. 20351. 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, Kory Kroft, Matthew Notowodigdo, 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.

alternative policy (Hicks (1940)). This means that everyone would prefer the transfers in the status quo to the alternative policy. Moreover, if the sum is positive, Kaldor (1939) argues that the winners could compensate the losers using individual-specific lump-sum transfers. Therefore, a modified policy that also includes compensating transfers would generate a Pareto improvement relative to the status quo. As a result, the Pareto principle implies the alternative should be preferred if and only if the sum of willingness to pay is positive. These two Pareto-based conceptual experiments underpin the standard definition of economic efficiency as the unweighted sum of individuals' willingnesses to pay.

Despite the aims of Kaldor and Hicks, measuring economic efficiency has become synonymous with a lack of consideration for interpersonal comparisons. Instead, researchers often employ a social welfare function to resolve the "equity-efficiency tradeoff". In doing so, one abandons the ability to provide normative guidance without relying on the subjective preferences of the researcher or policymaker.

This paper revisits these classic Kaldor-Hicks experiments. I argue that the criticism that Kaldor-Hicks does not account for equity concerns is related to the fact that the hypothetical policy changes in their conceptual experiments are infeasible. This is because the redistributive experiments envisioned by Kaldor and Hicks involve distortionary costs. Taxes are imposed on observable choices like incomes that respond to taxes and transfers (Mirrlees (1971)). Correctly measuring Kaldor and Hicks' original notion of economic efficiency requires accounting for these costs:

> *"Since almost every conceivable kind of compensation (re-arrangement of taxation, for example) must itself be expected to have some influence on production, the task of the welfare economist is not completed until he has envisaged the total effects...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).

The contribution of this paper is to provide a set of "efficient" welfare weights that provide a first-order implementation of the Kaldor and Hicks conceptual experiments using feasible modifications to the tax schedule. The efficient social welfare weight at an income level \$$y$ is equal to the marginal cost of providing \$1 of welfare to individuals earning near \$$y$. In a world without distortionary taxation, this cost is \$1. But, it differs from \$1 when individual choose to change their earnings in response to a \$1 tax cut to those earning near \$$y$. For example, those earning below \$$y$ might increase their incomes to \$$y$; those above \$$y$ may decrease their incomes to \$$y$. By the envelope theorem, those who change their earnings will not obtain a first-order utility improvement from the transfer; but, in the presence of taxes/transfers these responses have a first-order impact on government revenue. If taxes are positive (negative), increases in incomes create positive (negative) fiscal externalities. The efficient welfare weights differ from unity because of these fiscal externalities from redistributive taxes and transfers.

Weighting individuals' willingnesses to pay by efficient welfare weights implements the Kaldor and Hicks' tests of economic efficiency. If weighted surplus is negative, then modifications to the income tax schedule can lead to all points of the income distribution being better off than would be the case under the alternative environment. In this sense, the economist can suggest a better way to obtain the distribution of welfare gains offered by the alternative environment using a simple modification to the income tax schedule. Conversely, if the weighted surplus is positive, then a modified version of the alternative environment that redistributes the gains of the winners using feasible modifications to the tax schedule could make everyone better off. In this sense, the Pareto principle suggests preferring the alternative environment if and only if efficient-weighted surplus is positive.

What do the efficient welfare weights look like? First, I show that the weights can be expressed as a function of (a) the joint distribution of taxable income and marginal tax rates and (b) the behavioral response of taxable income to changes in taxation. I then use the universe of US income tax returns from 2012 to estimate this joint distribution, and I use it to provide bounds on the efficient welfare weights without assuming a magnitude of the behavioral response to taxation. To that aim, I show that the shape of the income distribution - in particular the local Pareto parameter of the income distribution - plays a key role in determining the extent to which the weights are above or below 1.[1] The estimation from the tax data shows that the Pareto parameter rises from near -1 at the bottom of the income distribution to near 2 at the top of the income distribution, crossing zero around the 60th quantile of the income distribution (around \$43K in ordinary income). As a result, weights are generally above one for those with incomes below \$43K, and below one for those with incomes above \$43K. Regardless of the size of the behavioral response to taxation, it is more costly to provide \$1 to the poor than to the rich. Thus, these bounds suggest it is efficient to weight surplus to the poor more than to the rich. Intuitively, it is more costly to move an additional \$1 from the top to the bottom of the income distribution through additional redistribution than it is to move \$1 from the bottom to the top of the income distribution through reduced redistribution.

Next, I construct point estimates of the efficient welfare weights using existing estimates of taxable income elasticities. The baseline specification suggests a \$1 tax cut to those with high incomes from a reduction in marginal tax rates costs around \$0.65. At the other end of the income distribution, the estimates suggest that expansions of the earned income tax credit (EITC) by \$1 to low earners has a fiscal cost of around \$1.15 because additional transfers cause individuals to adjust their earnings to maximize their tax credits. Combining, efficient welfare weights decline from around 1.15 at the bottom of the income distribution to around 0.65 at the top. In other words, \$1 to the poor can be turned into \$1.5-2 to the rich through modifications to the tax schedule. As a result, it is efficient to value surplus to the poor 1.5-2x more than surplus to the rich.

The weights can be used directly to compare feasible economic environments, such as different endowments or budget-neutral policy proposals. For example, the weights suggest an alternative environment that provides \$1.25 more to the rich but \$1 less to the poor would not be efficient. Even if one preferred \$1.25 to the high income person more than \$1 to a person at the bottom of the income distribution, it would be more efficient to provide this by modifying the tax schedule: this would generate \$1.5-\$2 in gains to the rich person for the same loss of \$1 to the poor person.[2] The efficient

[1] The Pareto parameter is given by $-\left(1+\frac{yf'(y)}{f(y)}\right)$ where $f(y)$ is the density of the income distribution.

[2] Following the spirit of Kaldor and Hicks' original desire to compare income distributions and measure societal welfare, Appendix F uses the efficient welfare weights to compare income distributions over time and across countries. I show that mean economic growth is significantly over-stated as a normative measure of changes in well being: growth would be 15-20% lower if taxation were used to hold fixed the changes in inequality over time. Comparing income distributions

welfare weights do not provide guidance on the exact amount of redistribution that is desirable.[3] Rather, their key benefit is to help researchers generate agreement on the most efficient method for redistribution.

In practice, many economic comparisons focus on expenditure policies that are not budget neutral. To that aim, I provide a welfare framework to evaluate whether a policy provides a potential Pareto improvement by constructing a policy's Marginal Value of Public Funds (MVPF). The MVPF is the ratio of the beneficiaries' willingness to pay for the policy change to the total cost of the policy change to the government inclusive of any fiscal externalities from behavioral responses to the policy (Hendren (2016)). I illustrate how to apply the framework by taking estimates of the MVPF from Hendren (2016) for job training programs, housing subsidies, and food stamps. This exercise shows how using the efficient welfare weights implements a comparison of these policies to a tax cut with the same distributional incidence. The resulting welfare framework does not rely on assumptions of the utility function (e.g. weak separability as in Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2008)), nor does it rely on a social welfare function. Instead, the normative conclusions rely solely on the positive analyses of economic policies combined with the Pareto principle.

Relation to Previous Literature This is not the first paper to recognize that incorporating the distortionary cost of transfers leads to a modification of the Kaldor and Hicks compensation principle. The theoretical analysis in this paper is most closely related to Coate (2000), who proposes an approach that incorporates the costs of redistribution into the Hicks criterion by comparing the policy to feasible alternatives such as distortionary redistribution through the tax schedule. Relative to this literature, the contribution of this paper is to provide a first-order implementation of this approach. [4]

The paper is also related to a literature deriving optimal policy conclusions without a social welfare function in cases when the utility function satisfies a weak separability assumption (e.g. Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2008)). When weak separability holds, the behavioral response to a policy in question is similar to the behavioral response to a distributionally-equivalent tax cut. But, many policies might violate weak separability (e.g. job training programs might complement labor supply by increasing human capital). The framework provided here does not require any assumptions about weak separability of the utility function.[5]

The derivation of efficient social welfare weights does not require the tax schedule to be at an optimum. But, if the tax schedule is set to maximize a social welfare function, the welfare weights are the solution to the "inverse optimum" program in optimal tax: they reveal the implicit preferences

across countries show that the income distributions of Austria and New Zealand would be preferred relative to the US income distribution, despite having a lower mean per capita income. Modifications to the tax schedule cannot replicate those distributions.

[3] Indeed, the welfare impact of changes to the progressivity of the tax schedule will always be zero because the weights implicily "undo" the change to the tax schedule.

[4] Coate (2000) writes: *"An interesting problem for further research would be to investigate whether the efficiency approach might be approximately decentralised via a system of shadow prices which convey the cost of redistributing between different types of citizens."* The efficient welfare weights are the shadow prices envisioned by Coate (2000).

[5] Appendix G shows how the welfare framework presented in Section 7 provides an empirical test of weak separability as assumed in Kaplow (1996, 2004, 2008) Intuitively, the test amounts to comparing the fiscal externalities from the policy in question to the fiscal externality of a distributionally-equivalent tax cut.

of those who are indifferent to modifications in the tax schedule (Bourguignon and Spadaro (2012); Zoutman et al. (2013, 2017)).[6] In this sense, the estimates suggest those indifferent to the current shape of the tax schedule in the US implicitly value an additional \$1 to the rich as equivalent to an additional \$1.5-2 to the poor.[7] The contribution of this paper is to show that these weights can be used to correctly implement the Kaldor-Hicks experiments, regardless of whether one believes the government set taxes optimally.

The methods in this paper are related to the large literature on optimal taxation (Mirrlees (1971); Saez (2001, 2002)). But, this paper has a very different aim. The efficient welfare weights generate indifference to modifications to the tax schedule. As a result, they do not provide guidance on the optimal degree of redistribution. Setting the optimal amount of redistribution in society still requires a social welfare function, and the approach here does not provide guidance on how to choose a social welfare function. Rather, they can help researchers understand whether food stamp expenditures might be more or less efficient than redistribution through other means such as the EITC. In this sense, efficient welfare weights implement the analyses originally envisioned by Kaldor and Hicks:

> *"All that economics can, and should, do in this field, is to show, given the pattern of income-distribution desired, which is the most convenient way of bringing it about?*
> *(Kaldor (1939, p552))*

While many may disagree about whether this is all that economics should do, this paper shows how to do Kaldor's desired analysis correctly by accounting for the distortionary cost of taxation and transfers.

The rest of this paper proceeds as follows. Section 2 derives the efficient welfare weights in the context of a general model setup. Section 3 illustrates how the weights implement the modified Kaldor-Hicks efficiency experiments. Section 4 provides a representation of these weights using the distribution of income, tax rates, and behavioral elasticities. Section 5 provides estimates of the joint distribution of income and tax rates and illustrates how they place bounds on the shape of the efficient welfare weights. Section 6 moves from bounds to point estimates by incorporating behavioral elasticities from previous literature. Section 7 illustrates how to apply these weights to study the efficiency of non-budget neutral policies using their marginal value of public funds. Section 8 concludes.

2 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 offered

[6]One exception to this invertibility is if the weights are negative. If the weights are negative, reducing taxes increases government revenue and corresponds to the efficiency test proposed by Werning (2007). In practice, plausible specifications suggest the weights are positive throughout the income distribution.

[7]The is related to concurrent work in Lockwood and Weinzierl (2016), who estimate the solution to the inverse optimum program in the U.S. using aggregated data from the Congressional Budget Office. In terms of interpreting the weights as the solution to the inverse optimum, the use of tax data in the present paper provides both bounds on these weights independent of the response to taxation and also a more precise estimate of the shape of these weights.

by an alternative environment and will also be used to describe the cost of redistribution through modifications in the tax and transfer schedule.

2.1 Setup

There exists a population of agents indexed by θ, with population size normalized to 1. There is a status quo environment and an alternative environment. In the status quo environment, agents consumption, c, and earnings, y. I allow each agent of type θ to have a potentially different utility function, $u(c, y; \theta)$, over consumption and earnings. I will not impose restrictions on the distribution of θ, so without loss of generality one can think of θ as simply indexing people in the population.

Agents choose c and y to maximize utility subject to a budget constraint,

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

where $T(y)$ is the taxes paid on earnings y and m is additional income beyond earnings.[8] With a slight abuse of notation, I let $c(\theta)$ and $y(\theta)$ denote the resulting choices of type θ.

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 value of m that is required for a type θ to obtain utility level v in the status quo environment.[9]

The goal is to evaluate the status quo relative to an alternative environment. To allow for a general set of possible alternative environments, let $u^a(c, y; \theta)$ denote the utility function for type θ in the alternative environment. Let $T^a(\circ)$ denote the tax schedule in the alternative environment so that the budget constraint is given by $c \leq y - T^a(y) + m$. Define $v^a(\theta)$ to be the level of utility obtained and $e^a(v; \theta)$ is the smallest value of m that is required for a type θ to obtain utility level v in the alternative environment.[10] Individual θ's willingness to pay (equivalent variation) for the alternative environment is then given by

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

This is the amount of additional money a type θ would need in the status quo to be as well off as in the alternative environment. In addition to this equivalent variation definition of willingness to pay, one could also construct a compensating variation measure using the expenditure function in the alternative environment, $cv(\theta) = e^a(v(\theta); \theta) - e^a(v^a(\theta); \theta)$. Because the distinction between equivalent and compensating variation is second order[11] and the approach below considers first-order adjustments, it will not be necessary to distinguish between equivalent or compensating variation in the analysis that follows.

[8]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 5.1.

[9]Formally, $e(v; \theta) = \inf\left\{m \mid \sup_{c,y}\{u(c, y; \theta) \mid c \leq y - T(y) + m\} \geq v\right\}$. Duality implies that $e\left(v^0(\theta); \theta\right) = m$.

[10]Analogous to the definition in the status quo environment, $e^a(v; \theta) = \inf\{m \mid \sup\{u^a(c, y; \theta) \mid c \leq y - T^a(y) + m\} \geq v\}$.

[11]For example, see Schlee (2013) for a recent discussion of the first-order equivalence of five common conceptualizations of willingness to pay, including compensating and equivalent variation.

I primarily consider alternative environments in which the willingness to pay for the alternative environment, $s(\theta)$, is homogeneous in income. With an abuse of notation, let $s(y)$ denote the willingness to pay for the alternative environment by a type θ who chooses income $y(\theta)$ in the status quo. When willingness to pay varies conditional on income, finding Pareto improvements using modifications to the tax schedule that transfers only based on observed income is of course more difficult. I return to a discussion of this in Section 3.4 and show that even in these more difficult cases the efficient welfare weights characterize potential Pareto comparisons.

Given $s(y)$, the goal is to answer two questions: (1) can the surplus, $s(y)$, can be replicated through modifications in the tax schedule in the status quo environment (i.e. the experiment in Hicks (1940))? And, (2) does there exists a modification to the tax schedule in the alternative environment that makes everyone better off relative to the status quo (i.e. the experiment in Kaldor (1939))? The answer to these questions will depend on how changes to the tax schedule affect government revenue.

2.2 Measuring Government Revenue

For any tax schedule, $\tilde{T}(\circ)$, let $R\left(\tilde{T}\right)$ denote government revenue in the status quo environment with tax schedule $\tilde{T}$,

$$R\left(\tilde{T}\right) = E\left[\tilde{T}\left(\tilde{y}\left(\theta; \tilde{T}\right)\right)\right]$$

where $\tilde{y}\left(\theta; \tilde{T}\right)$ denotes the earnings choice of a type θ when facing tax schedule, $\tilde{T}$, in the status quo environment.

I impose the regularity condition that tax revenue is continuously differentiable with respect to changes in the tax schedule. This would be immediately satisfied if individual behavioral responses were continuously differentiable (e.g. imposing standard quasi-convexity of the utility function). But, the regularity assumption allows for people to enter/exit the labor force, change jobs, or conduct other behavior that has discrete impacts on their earnings in response to tax changes. It only assumes these discrete responses average out across the population so that the magnitude of the aggregate behavioral response on tax revenue is smooth. Assumption 1 states this more formally.

Assumption 1. *For any function $h(y)$ of taxable income y, let $\tilde{T}_\epsilon(y) = T(y) + \epsilon h(y)$. Then R is continuously differentiable in ϵ for any function $h(y)$.*

In the alternative environment, let $R^a\left(\tilde{T}\right) = E\left[\tilde{T}\left(\tilde{y}^a\left(\theta; \tilde{T}\right)\right)\right]$ denote the revenue raised from a tax schedule $\tilde{T}(\circ)$, where $\tilde{y}^a\left(\theta; \tilde{T}\right)$ is the earnings choice of type θ in the alternative environment facing tax schedule $\tilde{T}(\circ)$. I assume Assumption 1 holds for R^a.

Importantly, note that Assumption 1 does not require continuity of behavioral responses to the policy. For example, it is easily satisfied in the economic environment of Zoutman et al. (2013) (and the environment we consider below) that allows for participation responses to labor supply.

2.3 Marginal Cost of Redistribution

Suppose individuals with income y^* are willing to pay $s(y) = \$1$ for the alternative environment. How much does it cost to replicate this \$1 benefit through a tax cut in the status quo environment? Figure 1 depicts a small tax deduction to those with 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 $y^* + \epsilon/2$.[12] By the envelope theorem, to first order individuals with earnings between $y^* - \epsilon/2$ and $y^* + \epsilon/2$ will be willing to pay η to have a tax schedule given by $\hat{T}(y; y^*, \epsilon, \eta)$ instead of $T(y)$.[13]

If there were no behavioral responses to changes in taxes, then the cost to the government of providing η to those earning near y^* would be η per person, so that the mechanical marginal cost in the absence of behavioral responses is 1. But, the presence of behavioral responses induce a fiscal externality on the government. To capture this, consider the derivative of revenue, $R\left(\hat{T}(\circ; y^*, \epsilon, \eta)\right)$, with respect to the size of the tax cut, η, and evaluate at $\eta = 0$. This yields the function $\frac{d\left[R\left(\hat{T}(\circ; y^*, \epsilon, \eta)\right)\right]}{d\eta}|_{\eta=0}$, 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^*:

$$g(y^*) \equiv \lim_{\epsilon \to 0} \frac{d\left[R\left(\hat{T}(\circ; y^*, \epsilon, \eta)\right)\right]}{d\eta}|_{\eta=0} \tag{2}$$

Equation (2) defines the efficient welfare weights. Since the choice of y^* was arbitrary, I use y to denote the argument of $g(y)$ instead of y^* going forward.

Definition 1. For each level of income, y, the **Efficient Welfare Weight**, $g(y)$, is the marginal cost of providing an additional \$1 to those earning near y, as defined in equation (2).

The cost of the tax cut to those earning near y is comprised of two components: a mechanical cost of \$1 and a fiscal externality, $FE(y) = g(y) - 1$

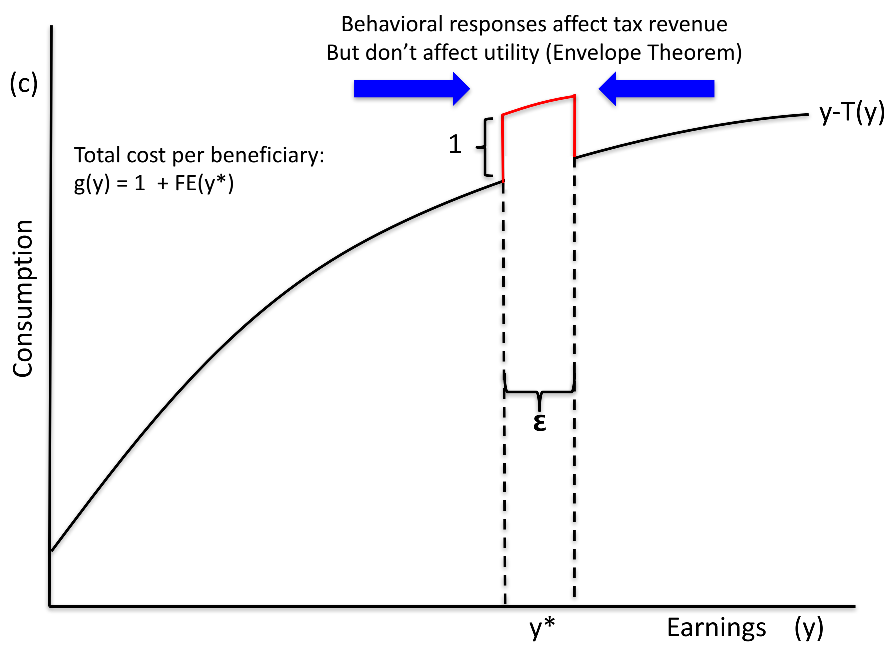
$$g(y) = 1 + FE(y)$$

[12] Note this is a discontinuous modification to the tax schedule. However, nothing in the analysis requires either T (or the modified schedule, $\hat{T}$) to be continuous or differentiable.

[13] This is true as long as the incidence of the tax cut falls entirely on the beneficiaries and does not result in changes in wages. For example, if firms respond to the tax cut of \$1 by lowering wages by \$0.50, then the individual would only be willing to pay \$0.50 for a \$1 tax cut. Here, I assume no general equilibrium responses, but this could be incorporated into future work. I discuss this further in Section 3.4.

If taxable income did not respond to changes in taxes, the marginal cost would be \$1 per beneficiary, $g(y) = 1$. The difference, $FE(y) = g(y) - 1$, equals the size of the fiscal externality from the behavioral response to the tax cut. The tax cut could cause people to work less and reduce tax revenue ($FE(y) > 0$); conversely, it could cause others to increase their tax payments, $FE(y) < 0$. The size of the fiscal externality is an empirical question: it depends on the causal impact of tax changes on the government budget. In Section 4, I provide additional assumptions that enable one to identify $FE(y)$ using behavioral elasticities, the shape of the income distribution, and the shape of the tax schedule.

Figure 1: Tax Cut to Those Earning Near y^*



Notes: This figure illustrates the modification to the tax schedule that provides a tax cut of \$1 to those with earnings in a region of y^* of width ϵ. To first order, those whose earnings would lie in $\left[y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right]$ will value the tax cut at \$1. But, the costs will result from both this mechanical cost and the impact of behavioral responses to the tax cut (loosely illustrated by the blue arrows). So, the total cost per unit of mechanical beneficiary will be $g(y) = 1 + FE(y)$.

The definition of $g(y)$ above is the marginal cost of providing an additional \$1 to those earning y in the status quo environment. For the alternative environment, I define $g^a(y)$ analogously.[14] Note that it may not be the case that $g^a(y) = g(y)$ since the alternative environment may have different tax schedules, distributions of income, and responses to taxation.

[14]Formally, let $\tilde{T}(\circ; y^*, \epsilon, \eta)$ denote a modified tax schedule in the alternative environment. Then,

$$g^a(y^*) \equiv \lim_{\epsilon \to 0} \frac{d\left[R\left(\tilde{T}(\circ; y^*, \epsilon, \eta)\right)\right]}{d\eta}|_{\eta=0}$$

is the marginal cost of providing additional resources to those with earnings y^* in the alternative environment.

3 Using the Efficient Welfare Weights

To set the stage, let $s(y)$ denote the willingness to pay for some alternative environment. Is the alternative environment preferred to the status quo? Before outlining the approach suggested here, first consider the social welfare function approach. This would weight willingness to pay by a generalized social welfare weight (Saez and Stantcheva (2016)) that balance the gains to the winners (e.g. for whom $s(y) > 0$) against the losses to the losers (e.g. for whom $s(y) < 0$). In contrast to this approach, I provide a motivation to weight this willingness to pay by the efficient welfare weights in equation (2), regardless of one's own social preferences.

I define the efficient surplus to be the amount of resources the government needs to replicate the surplus offered by the alternative environment using modifications to the tax schedule,

Definition 2. Efficient Surplus is given by a weighted average of surplus:

$$S = E\left[s(y)\, g(y)\right] \tag{3}$$

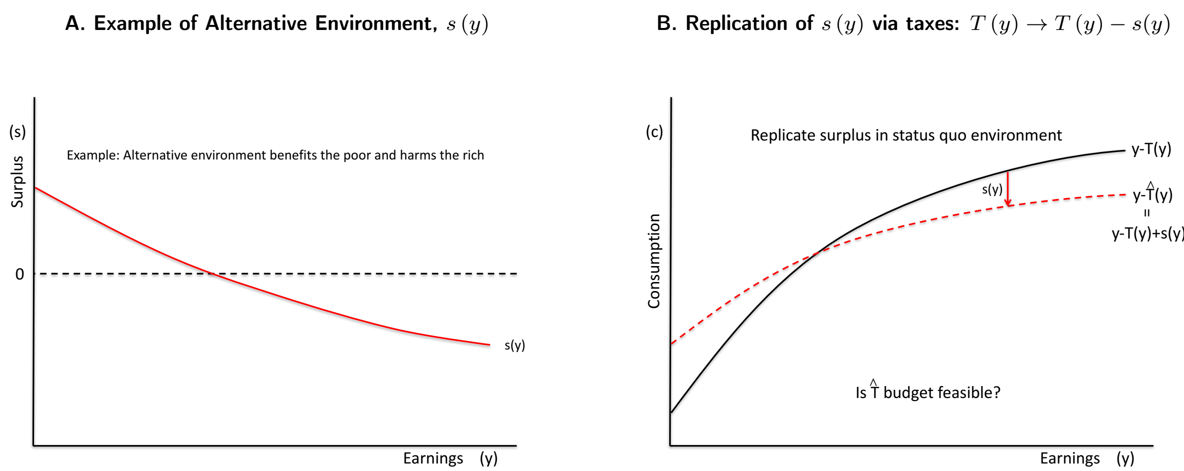
The next subsections illustrate how these weights and surplus measures implement the classic tests for efficiency in Kaldor and Hicks, modified to account for the distortionary cost of redistribution. Section 3.1 implements a first-order test of efficiency in the spirit of Hicks (1940): $S > 0$ if and only if one cannot replicate the surplus allocation offered by the alternative environment using modifications to the tax schedule. Section 3.2, implements a first order test of efficiency in the spirit of Kaldor (1939): $S > 0$ if and only if one can modify the tax schedule in the alternative environment to generate a Pareto superior allocation that is preferred by everyone relative to the status quo. Combined, these tests motivate testing for whether $S > 0$ to assess the desirability of the alternative environment.

3.1 Testing for Efficiency in the Spirit of Hicks (1940) and Coate (2000)

Can the benefits offered by the alternative environment, $s(y)$, be more efficiently provided through modifications in the tax schedule? To assess this, imagine replacing the current tax schedule, $T(y)$, with a new tax schedule, $\hat{T}(y) = T(y) - s(y)$, that offers a tax cut of size $s(y)$ to those earning y. Figure 2 provides an illustration. Panel A presents a hypothetical alternative environment that is preferred by the poor but not by the rich. Panel B then modifies the tax schedule from $T(y)$ to $T(y) - s(y)$. To first order, the envelope theorem implies that the tax cut of $s(y)$ is valued at $s(y)$ by those earning y. Therefore, everyone is approximately indifferent between the alternative environment and the status quo environment with the modified tax schedule, as depicted by the dashed red line in Figure 2, Panel B. The Hicks test for efficiency asks: Is this tax modification in the status quo world feasible?

To first order, the marginal cost of providing \$1 of welfare to those earning y is given by $g(y) = 1 + FE(y)$. Therefore, the cost of this tax cut is given by $E\left[g(y)\, s(y)\right]$. If this quantity is positive, then providing surplus $s(y)$ through the tax schedule would not be feasible. Closing the budget constraint by raising taxes on everyone would lead to the blue line in Figure 3, Panel A. In this sense, the alternative environment would be efficient relative to what is feasible through modifications to the tax schedule in the status quo. In contrast, if $S < 0$, then it is possible to replicate the alternative environment through modifications to the tax schedule. Redistributing the government surplus to everyone equally leads to the blue line in Figure 3, Panel B, which is preferred by all relative to the alternative environment. In this sense, the alternative environment is efficient if and only if $S > 0$.

Figure 2: Hicks (1940) Efficiency Experiment



Notes: This figure illustrates the efficiency experiment of Hicks (1940) for a hypothetical alternative environment. Panel A presents the hypothetical willingness to pay for each person at different points of the income distribution. In this example, those with low incomes prefer the alternative environment, but those with higher incomes prefer the status quo. Panel B illustrates modifying the tax schedule in the status quo world to attempt to replicate the surplus offered by the alternative environment. To first order, everyone is indifferent between the alternative environment and the modified status quo with tax schedule $T(y) - s(y)$.

The formal version of these statements are valid up to first order, as they rely on the envelope theorem to ensure indifference between the modified status quo (the dashed red line in Figure 2, Panel B) and the alternative environment. Proposition 1 provides one method of formalizing this idea by considering a scaled surplus function.

Proposition 1. *For any $\epsilon > 0$ define the scaled surplus by $s_\epsilon(y) = \epsilon s(y)$ and $S_\epsilon = E[s_\epsilon(y) g(y)] = \epsilon S$. If $S < 0$, 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, $s_\epsilon^t(y)$, that is uniformly greater than the surplus offered by the alternative environment, $s_\epsilon^t(y) > s_\epsilon(y)$ for all y. Conversely, if $S > 0$, no such $\tilde{\epsilon}$ exists.*

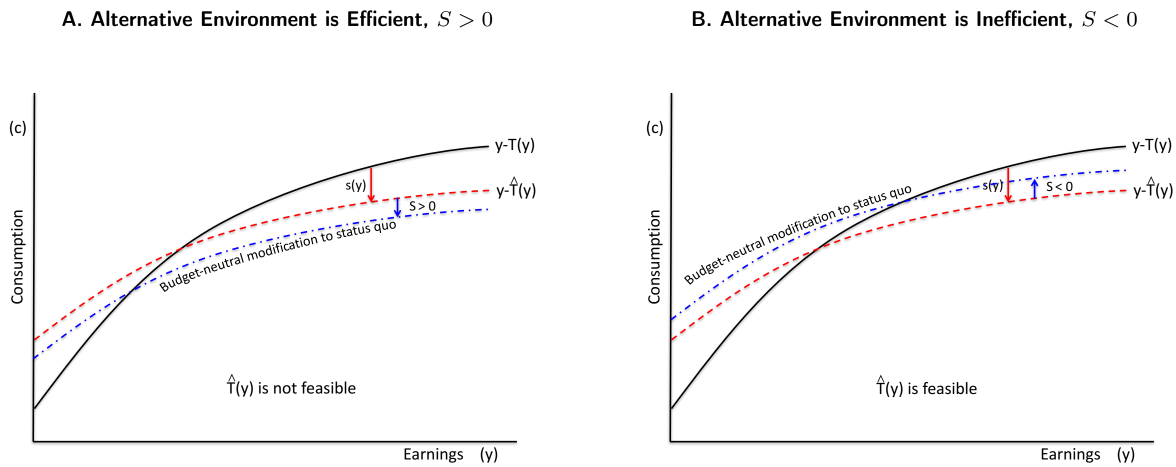
Proof. See Appendix A.2. □

Proposition 1 shows that the conclusions in Figures 2 and 3 hold to first order.[15] In this sense, the efficient welfare weights provide the right direction for adjusting for the marginal cost of redistribution.

[15] Proposition 1 formalizes the first order approach by scaling the surplus function. Alternatively, one could formalize the approach by directly modeling a continuum of alternative environments in the utility function. For example, suppose a is a continuous number indexing alternative environments (e.g. level of a public goods, trade policy, etc). Let $a = 0$ corresponds to the status quo and assume one can write individuals' utility functions, $u(c, y, a; \theta)$. In this case, one

However, the costs and benefits of redistribution through the tax schedule could differ for larger movements in the tax schedule. Moving beyond this first-order approach is a difficult but important direction for future work. In the meantime, testing $S > 0$ provides first-order guidance on how to correctly implement Hicks' original experiment in a way that accounts for the distortionary cost of redistribution.

Figure 3: Testing for Hicks Efficiency



Notes: This figure illustrates the efficiency test of Hicks (1939). The blue line illustrates the conceptual after-tax income that is feasible through modifications to the tax schedule but has the same distributional incidence as the alternative environment. Panel A illustrates the case in which the modified status quo tax schedule would deliver lower welfare to all points of the income distribution, so that the alternative environment is efficient relative to the status quo, $S > 0$. In contrast, Panel B illustrates the case in which replicating the surplus offered by the alternative environment through the tax schedule leads to higher welfare for all, so that the alternative environment is inefficient.

If $S < 0$, then the alternative environment is Pareto dominated by a modification to the tax schedule. In this sense, alternative environments for which $S < 0$ are not desirable. But, what about policies for which $S > 0$? Should these be pursued?

Armed with only the result in Proposition 1, it is unclear. While Hicks (1940) originally suggested yes, moving to the alternative environment does not generate a Pareto improvement relative to the status quo. Rather, it generates a Pareto improvement relative to a modified status quo that attempts to replicate the distributional incidence of the alternative environment. Actually moving to the alternative environment would generate winners and losers. Hence, $S > 0$ suggests it is a useful policy to consider (it's an "efficient" policy in the sense of Coate (2000)). But, it is not clear whether it is desirable relative to the status quo if $s(y) < 0$ for some y.

In order to provide guidance in the case when efficient surplus is positive, it is useful to consider a different conceptual experiment: that of Kaldor (1939).

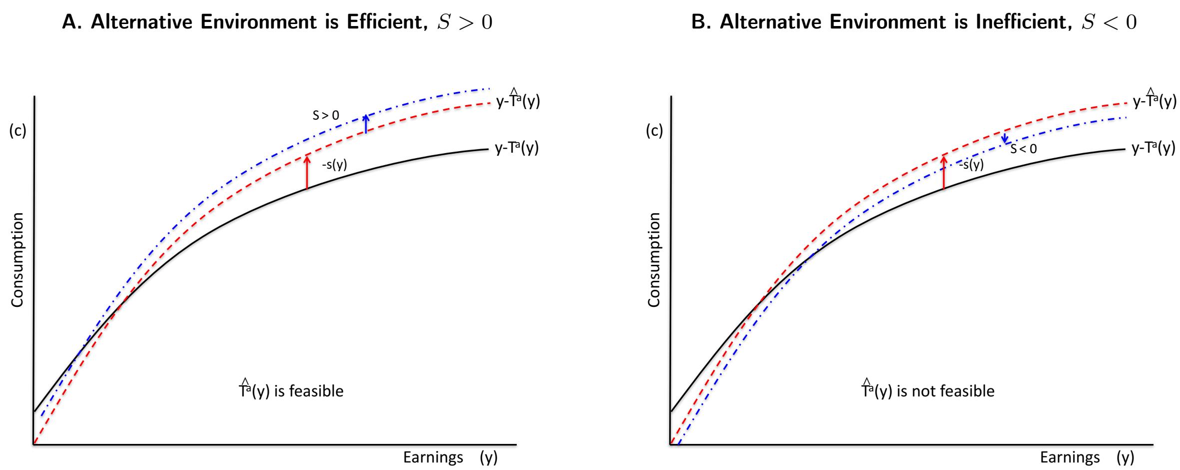
can define $s(y)$ to be individuals marginal willingness to pay out of their own income for a marginal change in a: $s(y) = \frac{\partial u}{\partial a} / \frac{\partial u}{\partial c}$ evaluated at $a = 0$. In this case, a modification to the tax schedule can make everyone better off relative to a world with a slightly higher value of a if and only if $E[g(y)s(y)] < 0$.

3.2 Finding Pareto Improvements in the Spirit of Kaldor (1939)

When can everyone be made better off relative to the status quo environment? Consider modifying the tax schedule in the alternative environment, $T^a(y)$, so that the winners compensate the losers, $T^a(y) \to T^a(y) + s(y)$.

Figure 4 presents this modified tax schedule in the alternative environment. Those with incomes y are better off by $s(y)$ relative to the status quo. The dashed red line in Figure 4 taxes back these gains. The envelope theorem suggests that to first order individuals earning y in the alternative environment are worse off by $s(y)$ when we tax back these benefits. Everyone is approximately indifferent between the status quo environment and the alternative environment with the modified income tax schedule. Therefore, the question becomes: Is this modification to the tax schedule in the alternative environment budget feasible?

Figure 4: Testing for (Kaldor) Efficiency



Notes: This figure illustrates the test of efficiency in Kaldor (1939) that modifies the tax schedule in the alternative environment to attempt to find a Pareto improvement in the modified alternative environment relative to the status quo. The dashed red line presents the after-tax schedule that adds the surplus offered by the alternative environment to the tax schedule, $T(y) + s(y)$. To first order, everyone is indifferent between the status quo and the modified alternative environment illustrated by the dashed red line in Panels A and B. The dash-dot blue line then illustrates the after tax income curve that results from closing the government budget constraint. Panel A illustrates the case that the alternative environment is efficient, so that after modifying the tax schedule in the alternative environment there is a Pareto improvement relative to the status quo. Panel B illustrates the case where the alternative environment is inefficient, so that after taxing back the benefits of the alternative environment and closing the budget constraint everyone is worse off relative to the status quo.

To first order, the modification to the tax schedule generates revenue $S^a = E[g^a(y)\, s(y)]$, where $g^a(y)$ is the cost to the government of providing \$1 to those earning near \$$y$ in the alternative environment. If $S^a > 0$, a modified alternative environment in which the winners compensate the losers through modifications to the tax schedule can make everyone better off relative to the status quo.

In practice, S^a could differ from S because the marginal cost of a tax cut may differ in the alternative and status quo environment, $g(y)$. If this is the case, it could be that the alternative environment dominates all feasible modifications to the status quo tax schedule ($S > 0$) but there does not exist a modified alternative environment that delivers a Pareto improvement relative to the status quo ($S^a < 0$). But, many applications involve sufficiently small changes to the structure of the economy, in which case it seems reasonable to assume that the marginal cost of taxation is similar in the status quo and alternative environments, $g^a(y) \approx g(y)$. I state this formally in Assumption 2.

Assumption 2. *For sufficiently small $\tilde{\epsilon}$, the marginal cost of taxation, $g(y)$, 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.*

If Assumption 2 holds, then $S > 0$ provides a first-order test of whether those with $s(y) > 0$ can compensate those with $s(y) < 0$ through modifications to the tax schedule in the alternative environment. Proposition 2 states this formally using the same scaled surplus function as in Proposition 1.

Proposition 2. *Suppose Assumption 2 holds. For $\epsilon > 0$, let $s_\epsilon = \epsilon s(y)$. If $S > 0$, 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 $s^t_\epsilon(y)$ that is positive at all points along the income distribution, $s^t_\epsilon(y) > 0$ for all y. Conversely, if $S < 0$, then no such $\tilde{\epsilon}$ exists.*

Proof. See Appendix A.3. □

In this sense, testing whether $S > 0$ provides a first-order approximation to searching for potential Pareto improvements as suggested by Kaldor.

Table 1

	S > 0	S < 0
Hicks Experiment: Possible to replicate s(y) using tax cut in status quo?	No	Yes
Kaldor Experiment: Possible to modify alternative environment tax schedule to make everyone better off relative to status quo?	Yes	No

Summary Table 1 summarizes the main results. When efficient surplus is negative, $S < 0$, the alternative environment is inefficient in the sense that a feasible modification to the tax schedule in the status quo environment can lead to a Pareto superior allocation to the alternative environment. In this sense, alternative environments for which $S < 0$ can be rejected by the logic of Hicks (1940) and Coate (2000).When efficient surplus is positive, $S > 0$, a modified alternative environment in which the winners compensate the losers through modifications to the tax schedule offers a Pareto superior allocation relative to the status quo. In this sense, the alternative can be preferred using the compensation principle in Kaldor (1939).

Of course, whether such Pareto comparisons are realized depends on whether the modifications to the tax schedule are actually implemented. But, in contrast to the traditional Kaldor-Hicks individual-specific lump-sum transfers, these modifications are feasible.

3.3 Relation to Previous Literature

3.3.1 Relation to Social Welfare Function and Inverse Optimum

Table 1 shows how efficient surplus, S, can evaluate policy changes by searching for potential Pareto improvements. In this sense one can make a recommendation about the alternative environment without resorting to a social preference that weighs the benefits to the winners against the losses to the losers. This contrasts with a more common approach to use a social welfare function to evaluate an alternative environment.

While the conceptual underpinnings in the Kaldor and Hicks experiments that rely on the Pareto principle differ from the social welfare function approach, the efficient welfare weights are related. The weights equal the implicit social welfare weights that rationalize indifference to modifications to the tax schedule. In other words, they correspond to the social preferences of someone who is indifferent to the shape of the current tax schedule.

Appendix B provides a formal derivation. To see the logic, let $\chi(y)$ denote the social marginal utilities of income for those earning near y, so that an additional \$1 to an individual earning y has an impact of $\chi(y)$ on social welfare. Suppose one provides a small tax cut of \$1 to those earning near y. Those with incomes near y will be willing to pay \$1 for this tax cut, and it will generate a social welfare impact of $1 * \chi(y)$. But, it will have a cost of $1 + FE(y) = g(y)$. Hence, the marginal value of additional government spending on a tax cut to those earning near y will be given by $\frac{\chi(y)}{g(y)}$. If the tax schedule is set to maximize social welfare, then the government must be indifferent between raising \$1 from those earning y' to finance a tax cut to those earning y. In other words, $\frac{\chi(y)}{g(y)}$ must be constant for all y;

$$\frac{\chi(y)}{g(y)} = \kappa \quad \forall y$$

So, $\chi(y) = \kappa g(y)$. Since social welfare weights are only defined up to a constant, $g(y)$ is the unique set of social welfare weights that rationalize the tax schedule as optimal.

In this sense, the efficient welfare weights are related to a literature estimating the inverse optimum program of optimal taxation. This literature estimates the implicit social welfare weights that rationalize the status quo tax schedule as optimal and uses them to infer properties of social or political preferences for redistribution (Bourguignon and Spadaro (2012); Zoutman et al. (2013, 2017)).

Relative to this literature, the contribution of this paper is to provide a justification to use the efficient welfare weights even if one is not indifferent to the shape of the current tax schedule. Regardless of one's own social preferences, and regardless of whether one believes the government has set the tax schedule optimally, the efficient welfare weights characterize the existence of potential Pareto improvements and implement the classic efficiency tests of Kaldor and Hicks.

3.3.2 Negative Weights and Werning (2007)

There is a special case when the weights, $g(y)$, identify a Pareto improvement directly through modifications to the tax schedule. In particular, if one finds that for some value of y the weights are negative, $g(y) < 0$, then the government can raise revenue by providing a tax cut to those earning near y. In this sense, testing whether $g(y) < 0$ implements a test in the spirit of Werning (2007) that searches for local Laffer effects in the tax schedule. A finding of $g(y) < 0$ makes a straightforward policy recommendation: cut taxes to those earning near y. In practice, the estimation will suggest $g(y) > 0$ for all y in the US for the range of plausible specifications considered.

3.4 Limitations

Efficient social welfare weights facilitate the search for Pareto comparisons. But, before turning to the empirical implementation and estimation, this subsection emphasizes several of the limitations of this approach.

First, the approach relies on the envelope theorem to measure the marginal cost of taxation. If a policy change is large enough, one could potentially be concerned that the modifications required to mimic the policies through the tax schedule may lead to a change in the marginal cost of taxation. For example, an earlier draft of this paper used these weights to compare the income distributions across countries. However, as will be clear below, the basic fact that the marginal cost of taxation is higher for the rich than the poor is quite robust because of the general shape of the income distribution. Thus, the efficient weights are more applicable to uses such as weighting the distributional incidence of food stamp expansions and other changes to expenditures on redistributive policies.

Second, the approach assumes that tax changes have no general equilibrium or spillover effects. Targeting a \$1 tax cut to those earning near \$$y$ is assumed to have a willingness to pay of \$1 for the beneficiaries of the tax cut. But, if their wages change in response to the tax cut, their willingness to pay may differ from \$1. Indeed, with spillovers and general equilibrium effects, the benefits of the tax cut may extend beyond those who are the direct target of the tax cut. But while taxation is not allowed to have GE effects, the approach does allow GE effects to drive the valuation of the alternative environment, $s(y)$. For example, the alternative environment could be a policy that makes more land available for agriculture, which in turn lowers food prices. One can still generate individuals' willingness to pay for this alternative environment, $s(y)$, and use the efficient welfare weights to ask whether this policy is efficient. In this sense, the efficient welfare weights, $g(y)$, are valid even if the policy change or alternative environment has GE effects; but it has ruled out the case where changes in the tax schedule, $T(y)$, has GE effects. I leave the incorporation of such effects for future work.

Indeed, recent work by Tsyvinski and Werquin (2018) provide one path forward using a structural model of taxation with GE effects.

Third, alternative environments may generate willingness to pay that is heterogeneous conditional on income. In this case, Pareto comparisons are more difficult. To test for Hicks efficiency, one needs to construct the maximum willingness to pay at each income level, $\overline{s}(y)$, and test whether $E[\overline{s}(y)\,g(y)] > 0$. If it is negative, then it would be feasible for the government to replicate the surplus offered by the alternative environment and make everyone better off. Intuitively, the government can feasibly provide a tax cut that covers even the maximal willingness to pay at each income level, $\overline{s}(y)$. In this sense, the alternative environment would be inefficient. Conversely, to test for Kaldor efficiency, one needs to construct the minimum willingness to pay at each income level, $\underline{s}(y)$, and test whether $E[\underline{s}(y)\,g(y)] > 0$. If it is positive, then it would be feasible for the government to redistribute income in the alternative environment so that everyone prefers the modified alternative environment relative to the status quo. Appendix C provides formal statements and proofs of these claims. Often, one might find that $E[\underline{s}(y)\,g(y)] < 0$ and $E[\overline{s}(y)\,g(y)] > 0$. In this instance, the alternative environment cannot not be Pareto-ranked relative to the status quo. Nonetheless, the efficient welfare weights, $g(y)$, continue to be the key component required to measure $E[\underline{s}(y)\,g(y)]$ and $E[\overline{s}(y)\,g(y)]$ that facilitates the search for these Pareto comparisons.

Lastly, it is important to emphasize that efficient welfare weights are not structural parameters. They are endogenous to the economic environment. There is no reason to expect efficient welfare weights identified in one setting or country to readily translate to another setting. This also has potentially interesting implications. For example, some have suggested that top tax rates in France are close to the top of the Laffer curve (Bourguignon and Spadaro (2012)), which implies that reductions in tax rates nearly pay for themselves, $FE(y) \approx -1$. In contrast, the point estimates below for the US will suggest smaller fiscal externalities. This would suggest testing for efficiency involves placing less weight on surplus accruing to the rich in France than in the US.

4 Representing Fiscal Externalities using Estimable Parameters

As illustrated in Figure 1, the marginal cost of providing a \$1 tax cut to those with earnings near y is given by $g(y) = 1 + FE(y)$, where $FE(y)$ is the impact of the behavioral response to the tax cut on government tax revenue. This section shows how these fiscal externalities can be written as a function of behavioral elasticities and the joint distribution of income and marginal tax rates. To do so, I impose the common assumption that taxation can cause individuals to enter (or leave) the labor force, but intensive margin adjustments are continuous in the tax rate. Assumption 3 states this more precisely in the context of the general model developed in Section 2.

Assumption 3. *Fix a type θ. For any $\kappa > 0$, let $B(\kappa) = [u(y(\theta) - T(y(\theta)), y(\theta); \theta) - \kappa, u(y(\theta) - T(y(\theta)), y(\theta);$ denote an interval of width κ near the status quo utility level. For any level of earnings y and utility* *level w, let $c(y;w,\theta)$ trace out a type θ's indifference curve that is defined implicitly by:*

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

I assume each 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 $c_{yy}>0$.*

Part (1) imposes the standard assumption that indifference curves vary smoothly with utility changes. Part (2) requires that indifference curves are convex on the region $y>0$ (but not at $y=0$). Importantly, it allows extensive margin responses: small changes in the tax schedule to cause jumps between $y=0$ and some positive income level. It is important to emphasize that Assumption 3 imposes very weak assumptions on utility functions and also allows for arbitrary distributions of unobserved heterogeneity, θ.[16]

When Assumption 3 holds, then three behavioral elasticities determine the response to taxation: a compensated elasticity, income elasticity, and participation elasticity. To define these, let $\tau(y)=T'(y)$ denote the marginal tax rate faced by an individual earning y. The average intensive margin compensated elasticity of earnings with respect to the marginal keep rate, $1-\tau(y)$, for those earning $y(\theta)=y$ is given by the percent change in earnings from a percent change in the price of consumption,

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

The average income elasticity of earnings, $\zeta(y)$, is given by the percentage response in earnings to a percent increase consumption,

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

The extensive margin (participation) elasticity with respect to net of tax earnings, $\epsilon^P(y)$, is given by

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

where $f(y)$ is the density of income at y.

Proposition 3 shows how these three elasticities along with the joint distribution of tax rates and income characterize the fiscal externality, $FE(y)$.

[16] See Kleven and Kreiner (2006) for a particular utility specification that satisfies Assumption 3 and captures these features of intensive and extensive margin labor supply responses.

Proposition 3. *For any point y such that $\tau(y)$ and $\epsilon^c(y)$ are constant in y and the distribution of y is continuous with density $f(y)$, the fiscal externality of providing additional resources to individuals near y is given by*

$$FE(y) = \underbrace{-\epsilon_c^P(y)\frac{T(y)-T(0)}{y-T(y)}}_{\text{Participation Effect}} \underbrace{-\zeta(y)\frac{\tau(y)}{1-\frac{T(y)}{y}}}_{\text{Income Effect}} \underbrace{-\epsilon^c(y)\frac{\tau(y)}{1-\tau(y)}\alpha(y)}_{\text{Substitution Effect}} \tag{4}$$

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

Proof. Proof provided in Appendix A The appendix also provides a generalized formula for points y such that $\epsilon^c(y)$ is not constant in y.[17] □

The fiscal externality associated with providing an additional dollar resources to an individual earning y is the sum of three effects. First, people may enter the labor force. $\epsilon^c(y)$ measures the size of this effect. It's impact on tax revenue depends on the difference between the average taxes received at y, $T(y)$, 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. The size of this effect is measured by $\zeta(y)$. The impact of this change in earnings on the government budget depends on the marginal tax rate, $\tau(y)$.

Finally, people earning close to y may change their earnings towards y in order to get the transfer. The elasticity, $\epsilon^c(y)$, measures how much people move their earnings towards y in response to the tax cut. The tax ratio, $\frac{\tau(y)}{1-\tau(y)}$, captures the fiscal impact of these responses. However, the net impact on government revenue is the sum of two effects. Some people will decrease their earnings towards y; others will increase their earnings towards y, as depicted by the blue arrows in Figure 1. When $\tau(y) > 0$, the former effect increases tax revenue and the latter effect decreases tax revenue. The extent to which the losses outweigh the gains depends on the elasticity of the income distribution, $\frac{yf'(y)}{f(y)}$. When $\frac{yf'(y)}{f(y)} < -1$ (as is the case with the Pareto upper tails in the US income distribution), more people increase rather than decrease their taxable earnings. This means $\alpha(y) > 0$. Conversely, if $\frac{yf'(y)}{f(y)} > -1$ (e.g. if f is a uniform distribution so that $f'(y) = 0$), then more people decrease than increase their earnings so that $\alpha(y) < 0$. This increases the marginal cost of the tax cut. Importantly, this shows that 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 and the shape of efficient welfare weights.

[17] Proposition 3 is a generalization of the canonical optimal tax formula to the case of multi-dimensional heterogeneity. Consistent with the intuition provided by Saez (2001), Proposition 3 shows that the relevant empirical elasticities in the case of potentially multi-dimensional heterogeneity are the population average elasticities conditional on income. Most closely related to this derivation, Zoutman et al. (2017) provide a derivation of equation 4 for the case when θ is uni-dimensional and distributed over an interval and for which the Spence-Mirrlees single crossing condition holds.

5 Bounds on Efficient Welfare Weights in the U.S.

At first glance, equation (4) suggests one requires precise estimates of the size of behavioral responses to taxation in order to quantify the efficient welfare weights. This is potentially problematic because of the general lack of consensus on the size of behavioral responses to taxation (Saez et al. (2012)). Fortunately, under fairly plausible assumptions outlined below, the shape of the income distribution provides insights into the shape of the efficient welfare weights. In particular, it will suggest the weights are generally higher for those with low incomes than high incomes.

5.1 Sample and Variables

To estimate the joint distribution of income and tax rates, I use the universe of de-identified 2012 tax returns taken from the 2012 IRS-SOI Databank maintained under the Statistics of Income Division at the IRS. I focus on primary filers aged 25-60 and their married spouses, if applicable.[18] Following Chetty et al. (2014), I restrict the sample to households with positive family income. Details of the sample and data construction are provided in Appendix D; Appendix Table I presents the summary statistics. The resulting sample has roughly 100 million filing units.

I define y to be the tax filer's ordinary income in 2012.[19] This 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 subtracts deductions for things like the number of children and charitable donations.

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 on the EITC schedule using information on the number of EITC-eligible children (reported 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.[20] In the end,

[18] 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.

[19] Because ordinary income determines the federal tax, it is the notion of income that most closely aligns with the theory.

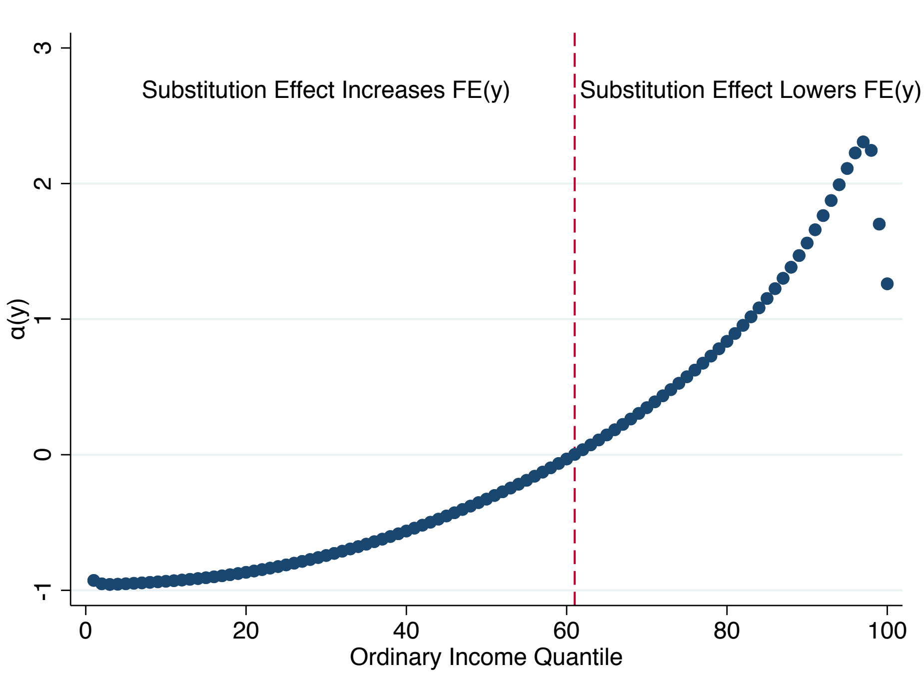
[20] Choosing alternative values for the state tax rates or the EITC rates do not significantly alter the results; as discussed below, the primary primary driver of the shape of the weights is the Pareto parameter combined with the assumption of a constant elasticity, not the shape of tax rates, $\tau(y)$.

this generates a marginal tax rate, $\tau(y)$, faced by each filer on an additional dollar of income.

5.2 Estimation of Shape of Income Distribution, $\alpha(y)$

I estimate $\alpha(y)$ for each level of income separately for those facing different tax schedules. The key additional complication that one needs to address is how to deal with the fact that two tax filers with the same ordinary income may face different tax schedules.[21] With multiple tax schedules, the average fiscal externality at each level of income requires estimating the average value of the substitution effect, $\epsilon\frac{\tau(y)}{1-\tau(y)}\alpha(y)$, separately for each tax schedule, $\tau(y)$, and then averaging across all individuals at each income level, y. I construct this by separately estimating $\alpha(y)$ for each tax schedule using the information in the tax returns on filing status and other determinants of the tax schedule. The details of this procedure are provided in Appendix D.

Figure 5: Shape of the Income Distribution, $\alpha(y)$



Notes: This figure presents the estimates of the average $\alpha(y)$ for each quantile of the income distribution. This function is given by $\alpha(y) = -\left(1 + \frac{yf'(y)}{f(y)}\right)$, where $f(y)$ is the density of the income distribution. For values of y below the 60th quantile, $\alpha(y) < 0$ so that the substitution effect in equation (4) raises the marginal cost of taxation. In contrast, for values of y above the 61st quantile, $\alpha(y) > 0$ so that the substitution effect lowers the marginal cost of taxation.

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

5.3 Results

Figure 5 presents the mean value of $\alpha(y)$ at each quantile of the ordinary income distribution. The average $\alpha(y)$ reaches around 1.5 at the top of the income distribution, consistent with findings in previous literature focusing on top incomes (Diamond and Saez (2011) and Piketty and Saez (2013)). However, the key point on Figure 5 is that $\alpha(y)$ exhibits considerable heterogeneity across the income distribution. It is negative below the 60th percentile of the income distribution, $\frac{yf'(y)}{f(y)} > -1$. This implies that the substitution effect increases the marginal cost of a tax cut (assuming a positive elasticity).[22] Conversely, it crosses zero around the 60th percentile, and is then positive. This means that $\frac{yf'(y)}{f(y)} < -1$ for values of y above the 60th quantile. For those earning more than about \$43K in ordinary income, the substitution effect reduces the cost of providing a tax cut. As long as $\tau(y) > 0$ and $\epsilon^c(y) > 0$, the substitution effect, $-\epsilon^c(y)\frac{\tau(y)}{1-\tau(y)}\alpha(y)$, in equation (4) is positive for incomes below \$43K (60th quantile of 2012 ordinary income) and negative for incomes above \$43K.

In addition to the substitution effect, it is also possible to put bounds on the natural shape of the impact of the participation effect on the government budget. For those with low incomes, the EITC offers transfers for those who enter the labor force; this renders $T(y) < 0$ so that those who enter the labor force in response to an increased tax cut actually increase the budgetary cost because they obtain the EITC benefits. In contrast, for higher values of y individuals contribute positive tax revenue so that $T(y) > 0$; thus any increase in labor force participation for those at higher income levels will result in a positive fiscal externality. This suggests the participation effect in equation (4) is also declining in y.

Lastly, most empirical works suggests income effects effects are either small (Gruber and Saez (2002); Saez et al. (2012)) or declining in income (Cesarini et al. (2015)). As a result, one has a natural bound on the shape of efficient welfare weights: Efficient welfare weights put greater weight on those with lower incomes (i.e. below \$43K) than those with higher incomes (i.e. above \$43K). This means that it is costly to redistribute an additional dollar from rich to poor, but cheap to redistribute from poor to rich.

6 Quantifying Efficient Welfare Weights for the U.S.

6.1 Elasticity Calibration

Point estimates of $g(y)$ require estimates of the behavioral responses to taxation. For those subject to the EITC, I draw upon Chetty et al. (2013) who calculate 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)).

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

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. I assess the robustness to alternative elasticities such as 0.1 and 0.5.

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.[23] Therefore, I assume the participation effect in equation (4) 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.

Lastly, 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)). Cesarini et al. (2015) find evidence of income effects using Swedish lotteries; however a large portion of these effects are driven by extensive margin responses and arguably already captured by the EITC responses measured above.[24]

6.2 Results

I use equation (4) to combine the estimates of the shape of the income distribution, marginal tax rates, and elasticity calibrations, which generates an estimate of $FE(y)$ for each filer. I then bin the income distribution into 100 quantile bins and construct the mean fiscal externality, $FE(y)$, for each quantile of income. The efficient social welfare weight at each income quantile is then given by $g(y) = 1 + FE(y)$. Figure 6 presents the resulting estimates for $g(y)$. Figure 7 presents the results for the alternative calibrations of the compensated elasticity of $\epsilon^c = 0.1$ and $\epsilon^c = 0.5$.

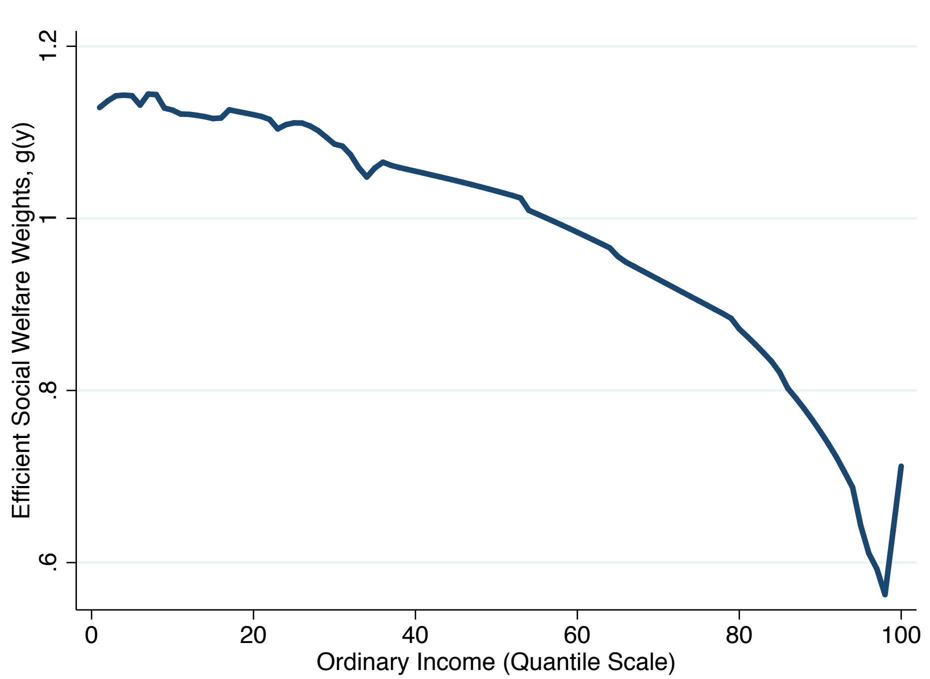
The weights have several key features. First, consistent with the bounds shown in the previous section, the results suggest it is efficient to place higher weight on surplus to the poor than to the rich. Under the baseline specification, these weights fall from around 1.15 for those at the bottom of the income distribution to 0.65 for those at the top. Transferring \$1 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. Conversely, transferring \$1 from the bottom of the income distribution can generate around $1.15/0.65 = \$1.77$ of welfare to the those at the top of the income distribution.

Second, although the weights place more weight on low versus high income individuals, the weights never differ by more than a factor of 2. In other words, $\left|\frac{g(y)}{g(y')}\right| < 2$ for all y and y'. This means that it is

[23] See Hotz and Scholz (2003) for a summary of elasticities and Hendren (2016) for the 9% calculation.

[24] Nonetheless, Appendix E reports the robustness of the results to an alternative specification that incorporates income effects assuming that the estimates from Cesarini et al. (2015) are entirely along the intensive margin and correspond to an elasticity of $\zeta = 0.15$. As shown in Appendix Figure 3, income effects tend to increase the marginal cost of taxation at all income levels; but in contrast to the compensated elasticity they do not affect the relative difference in the weights to low versus high income individuals.

Figure 6: Efficient Welfare Weights, $g(y)$



Notes: This figure presents the baseline estimates of the efficient welfare weights, $g(y)$, estimated using equation (4) for each quantile of the income distribution.

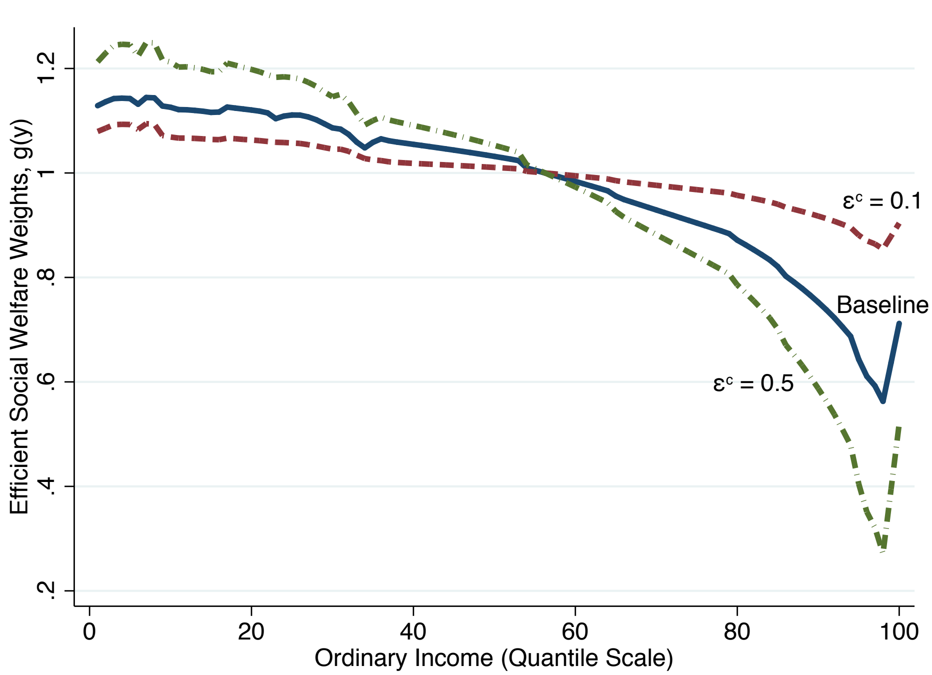
not efficient to discount surplus more than 50%, regardless of where it falls in the income distribution. For example, the consumer surplus standard in merger analysis sums only the willingness to pay by the consumers, and applies zero weight to producers or profits. The analysis here shows that at most one would want to down-weight surplus to the rich or to producers by a factor of 2. Applying no weight to producer surplus would still not be efficient even after accounting for the distortionary cost of taxation.

Third, while the weights generally decline in income, there is an increase in the top 1%. For the baseline specification, it is cheaper to provide additional transfers to the upper middle class than to the top 1%. However, Figure 7 illustrates that this non-monotonicity is not robust to plausible assumptions about how elasticities might change across the income distribution. In particular, if the elasticity moves from 0.3 to 0.5 as one goes from the top 2% to the top 1%, the weights would again be monotonically declining in income.

Fourth, all the weights are positive, $g(y) > 0$ for all y for the baseline and alternative specifications. This means that it is always costly to provide a tax cut. This implements a Pareto efficiency test suggested by Werning (2007), and suggests there are no Pareto improvements solely from modifying the tax schedule.

Lastly, as foreshadowed by the bounding exercise in the previous Section, there is a similarity between the estimates of $\alpha(y)$ in Figure 5 and the shape of the efficient welfare weights, $g(y)$. Higher elasticities, $\epsilon^c(y)$, increase the difference between the weights on the low- versus high-income individ-

Figure 7: Robustness to Alternative Elasticities



Notes: This figure presents the baseline specification for the efficient welfare weights alongside with estimates under alternative constant compensated elasticity scenarios of $\epsilon^c(y) = 0.1$ and $\epsilon^c(y) = 0.5$.

uals. But, they do not affect the general conclusion that $g(y) > 1$ for those with low incomes and $g(y) < 1$ for those with high incomes.

Applications The efficient welfare weights can compare economic environments, such as two feasible endowments or conceptual comparisons of income distributions. Indeed, part of the aim of the original Kaldor and Hicks papers was to measure societal well-being using data on income distributions. To that aim, Appendix F illustrates this application in detail by comparing income distributions across countries and over time.

As is well-known, the US has greater income inequality than other countries. Figure 6 shows that it would be costly to reduce this inequality, as it would require redistribution from rich to poor. Consistent with this, the results in Appendix F show that this can reverse some comparisons of income distributions across countries: although the US has higher mean per capita income than Austria and New Zealand, if the US tried to replicate their income distributions, it would make everyone worse off because of the distortionary cost of taxation.

Similarly, income inequality has increased over time in the U.S., especially at the upper regions of the income distribution. If the tax schedule was used to redistribute this increased inequality equally across the income distribution, the change in well-being since 1980 would be 15-20% lower than is implied by the increase in mean incomes. Of course, these comparisons are primarily illustrative – one should question the validity of invoking the envelope theorem to differences in the income distributions across countries or changes in the US over the last 30 years. But, they nonetheless highlight the potential importance of accounting for distributional incidence when conducting comparisons of income distributions in the spirit of the original Kaldor and Hicks ideas.

7 Applying the Weights to Analysis of Non-Budget Neutral Policy Changes

The efficient welfare weights can readily be applied to compare the status quo to a feasible alternative. But, often one wishes to evaluate expenditure policies of the government. Should we increase food stamp spending? Reduce Medicaid spending? Provide free public transportation? In these cases, willingness to pay may be positive for everyone, but one also needs a method to account for the cost to the government of the policy.

Traditional approaches to providing such guidance generally require a social welfare function or assumptions on the utility function, such as weak separability (Hylland and Zeckhauser (1979); Kaplow (1996, 2004, 2008)). Here, I show how the efficient welfare weights yield a welfare framework that does not require assumptions on utility functions, nor does it require a social welfare function. Instead, one can use the efficient welfare weights to implicitly compare the policy to a tax cut with similar distributional incidence.

7.1 Setup

To fix intuition, consider a policy that affects those with income near $\bar{y}$ (the more general case will be considered below). Assume the researcher has measured both (a) individuals' willingnesses to pay for the policy change, $\bar{s}$, and (b) the net cost of the policy per beneficiary, c. Importantly, c should incorporate any fiscal externalities from the policy change. For example, if the policy builds roads that increase labor earnings, it should incorporate the resulting increase in tax revenue.

Now, consider the Hicks' experiment in which the government tries to replicate $\bar{s}$ through modifications to the tax schedule. This would cost $\bar{s}g(\bar{y})$. It would be cheaper to replicate this surplus through the tax schedule if and only if

$$\bar{s}g(\bar{y}) \geq c \tag{5}$$

If equation (5) holds, it is more efficient to provide a tax cut to those earning near $\bar{y}$ than it is to increase spending on the policy. Note equation (5) can be extended to the case where there are multiple beneficiaries with willingness to pay $s(y)$. In this case, the LHS of equation (5) would be $E[s(y)g(y)]$, instead of $\bar{s}g(\bar{y})$.[25]

[25] The bias from using $\bar{s}g(\bar{y})$ instead of $E[g(y)s(y)|y \in Y]$ comes from two sources: nonlinearities in $g(y)$ and covariance between $g(y)$ and $s(y)$ amongst the beneficiaries,

$$E[g(y)s(y)|y \in Y] = g(\bar{y})\bar{s} + \underbrace{(E[g(y)|y \in Y] - g(\bar{y}))\bar{s}}_{\text{Nonlinearity in } g(y)} + \underbrace{cov(g(y), s(y)|y \in Y)}_{\text{Cov of WTP with } g(y)}$$

These biases are small when the income of the target population is concentrated around a particular $\bar{y}$ or if the efficient

Conversely, one can consider the Kaldor experiment: redistribute the benefits of the policy to attempt to create a Pareto improvement. Taxing back the benefits lowers taxes by $-\bar{s}$ to those earning near $\bar{y}$. Assuming $g(y)$ continues to measure the marginal cost of taxation (i.e. Assumption 2), this generates a cost to the government of $g(\bar{y})(-\bar{s})$. So, the total cost of the combined policy is $-g(\bar{y})\bar{s} + c$. If equation (5) holds, then this total cost is negative. This means there are sufficient resources for the winners to compensate the losers through modifications to the income tax schedule. In this sense, equation (5) characterizes whether increased spending on the policy is efficient.

The intuition of these comparisons is straightforward: compare the marginal value of public funds (MVPF) of the policy to the MVPF of a tax cut with similar distributional incidence. The MVPF of a policy is defined as the beneficiaries willingness to pay divided by the net cost to the government, $MVPF = \frac{\bar{s}}{c}$.[26]

If the MVPF of the policy is $\frac{\bar{s}}{c}$, what is the MVPF of the tax cut? The benefit of the \$1 tax cut to the targeted beneficiaries is \$1. The cost to the government of this targeted tax cut is precisely the definition of the efficient welfare weight, $g(y)$. So, the MVPF of a tax cut targeted to those with incomes near $\bar{y}$ is $\frac{1}{g(\bar{y})}$. Re-writing equation (5) yields

$$MVPF = \frac{\bar{s}}{c} \geq \frac{1}{g(\bar{y})} \tag{6}$$

The Kaldor and Hicks experiments suggest comparing the MVPF of the policy, $\frac{\bar{s}}{c}$, to the MVPF of a tax cut with the same distributional incidence, $\frac{1}{g(\bar{y})}$.

Relation to "The" Marginal Cost of Funds A more common approach in previous literature to evaluating economic policies is to ask whether willingness to pay, $\bar{s}$, exceeds $(1+\phi)c$, where $1+\phi$ is "the" marginal cost of public funds. Heuristically, ϕ is intended to account for the distortionary cost of taxation, generally assumed to be 0.3 or 0.5. Importantly, the same ϕ is applied regardless of whether the policy targets high or low income beneficiaries.

Equation (6) shows that applying a uniform cost ϕ does not correspond to testing for efficiency. If it is costly for society to redistribute to the poor (e.g. $g(\bar{y}) > 1$), then the relevant benchmark for $\frac{\bar{s}}{c}$ is not 1, 1.3, or 1.5, but rather a number less than 1 that corresponds to the cost to the government of providing welfare to the poor. Conversely, it is cheap to redistribute additional resources to the rich, $g(\bar{y}) < 1$. Therefore a policy must offer significant benefits per dollar of government expenditure to be more efficient than what can be offered through a tax cut to high-income individuals. The Kaldor-Hicks logic motivates a simple test: ask whether the policy is a more efficient method of redistribution than modifications to the tax schedule with similar distributional incidence.

welfare weights are relatively constant within the beneficiary population.

[26] An MVPF of $\frac{\bar{s}}{c}$ means that \$c of net government expenditure on the policy generates $\bar{s}$ units of welfare to the beneficiaries in units of their own willingness to pay. See Hendren (2016) for a detailed discussion of the MVPF and its properties. Although sometimes labelled the marginal excess burden and marginal cost of public funds (which are both conceptually distinct), it was originally mathematically formulated in Mayshar (1990) and subsequently analyzed in Slemrod and Yitzhaki (1996) and Kleven and Kreiner (2006).

Relation Weak Separability Some influential theoretical work suggests that even in the presence of distortionary taxation, it can be optimal to simply compare individuals' (unweighted) willingness to pay to the mechanical cost of the policy, excluding any fiscal externalities generated by the policy (Hylland and Zeckhauser (1979); Kaplow (1996, 2004, 2008)). These results rely on the assumption that utility satisfies weak separability.[27] Appendix G shows that the framework developed here nests these theoretical results and provides a test for the validity of the weak separability assumption. Here, I provide the intuition for the particular case when the policy only affects those with incomes $\bar{y}$.

Suppose we increasing spending by \$1 on a policy and let $c = 1 + FE_G$ denote the net cost to the government. Here, $FE_G = c - 1$ is the impact of the behavioral response to the policy on the government budget (i.e. its fiscal externality). Similarly, recall that one can write $g(\bar{y}) = 1 + FE(\bar{y})$, where $FE(\bar{y})$ is the fiscal externality from the tax cut targeted to those with earnings near $\bar{y}$.

With this notation, equation (6) can be re-written as preferring the policy if and only if:

$$\bar{s}(1 + FE(\bar{y})) \geq 1 + FE_G$$

or

$$\bar{s} - 1 \geq FE_G - \bar{s}FE(\bar{y}) \tag{7}$$

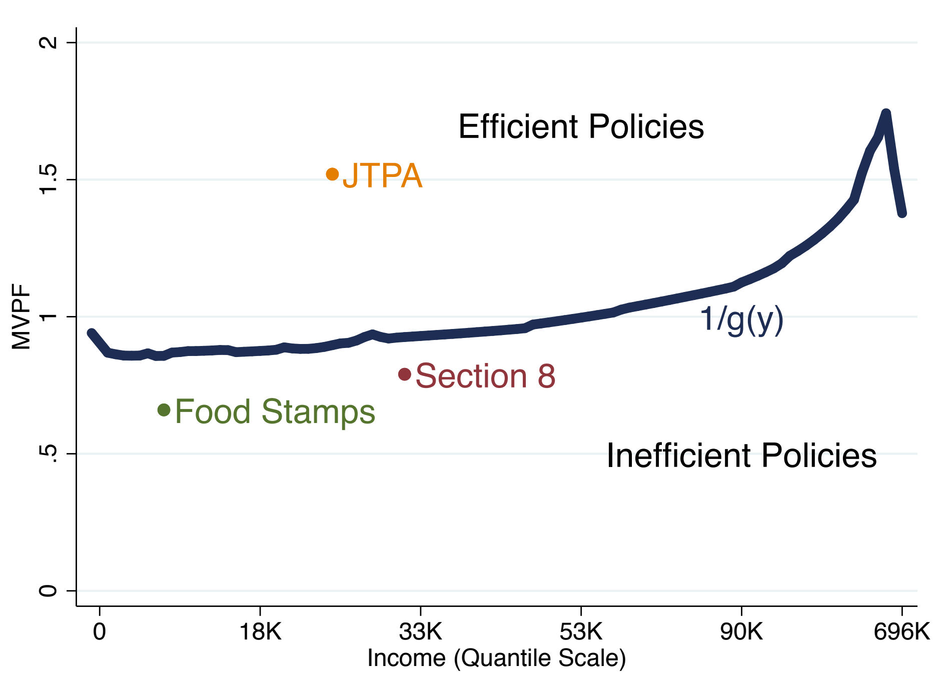
The LHS of equation (7), $\bar{s} - 1$, the standard "Samuelson rule" that compare willingness to pay, $\bar{s}$, to the \$1 mechanical cost of the policy. The RHS is the difference between the fiscal externality from the policy, FE_G, and the fiscal externality of a tax cut scaled by the willingness to pay for the policy, $\bar{s}$. It turns out that weak separability of the utility function implies $FE_G = \bar{s}FE(\bar{y})$. This is because, when weak separability holds, the behavioral response of taxable income to a tax cut of size $\bar{s}$ is equal to the behavioral response of taxable income to the policy G. As a result, these two fiscal externalities cancel out so that equation (6) reduces to testing whether $\bar{s}$ exceeds the mechanical cost of the policy before incorporating any of its fiscal externalities. This then reduces to the standard Samuelson rule of testing whether willingness to pay exceeds the mechanical cost, as Hylland and Zeckhauser (1979); Kaplow (1996, 2004, 2008). In contrast, equation (6) provides a test of potential Pareto improvements without imposing assumptions on the utility function or behavioral responses.

7.2 Illustrative Application

I illustrate the welfare framework by studying the efficiency of three policies whose MVPFs are computed in Hendren (2016). These policies include Section 8 housing vouchers, food stamps, and job training. Hendren (2016) provides details on the construction of the MVPF for each of these policies. The MVPF for Section 8 vouchers draws upon estimates of the fiscal externality inferred from the causal effects in Jacob and Ludwig (2012) on adult labor supply. The MVPF for food stamps draws upon estimates of the labor supply fiscal externalities from Hoynes and Schanzenbach (2012). And, the MVPF for job training draws upon estimates of the impact of the Job Training Partnership Act

[27] In this context, utility function, $u(c, y, G; \theta)$, is weakly separable if it can be written as $u(v(c, G), y; \theta)$ where $v(c, G)$ is a function that is common to all individuals, θ.

Figure 8: Illustrating the Test for Efficiency of Policy Changes



Notes: This figure illustrates the use of the efficient welfare weights for assessing the efficiency of government policy changes. The line presents the value of $\frac{1}{g(y)}$, which represents the amount of welfare that can be delivered to each portion of the income distribution per dollar of government spending. The dots present estimates of the marginal value of public funds (MVPF) for three policy examples: the job training partnership act (JTPA) from Bloom et al. (1997), food stamps from Hoynes and Schanzenbach (2012), and Section 8 housing vouchers from Jacob and Ludwig (2012). The vertical axis presents the estimated MVPF from Table 1 of Hendren (2016); the horizontal axis presents the estimated income quantiles of the beneficiaries of each policy (normalized to 2012 income using the CPI-U). An MVPF that falls above (below) the the Income/EITC line correspond to policies that can(not) generate Pareto improvements.

(JTPA) on labor earnings and benefit substitution from Bloom et al. (1997).

To illustrate the construction of the MVPF, consider the food stamp application. Hoynes and Schanzenbach (2012) estimate the program cost of the food stamp program was $1,153 per household. But, reductions in labor supply led to reduced federal and state income taxes of $588. Hence, the net cost of the program was $1,741. In contrast, assuming the food stamps are primarily infra-marginal to the budget of the beneficiaries (consistent with Whitmore (2002)), the envelope theorem suggests beneficiaries value food stamps at their program cost of $1,153. Hence, the MVPF of the food stamp program is given by

$$MVPF^{FoodStamps} = \frac{\bar{s}}{c} = \frac{1,153}{1,153+588} = 0.66 \tag{8}$$

As is noted in Hendren (2016), there is considerable statistical and model uncertainty around each of the point estimates of the MVPF for each of these policies. To emphasize the illustrative nature of this exercise, note that the estimate of the fiscal impact of $588 from Hoynes and Schanzenbach (2012) is not statistically distinguishable from zero (which means that the MVPF of food stamps is not statistically distinguishable from 1).

Figure 8 plots the $MVPF = \frac{\bar{s}}{c}$ for each of these policies taken directly from Table 1 of Hendren (2016). The horizontal axis corresponds to the quantile corresponding to the mean income, $\bar{y}$, of the policy beneficiaries.[28] The point estimates suggest that food stamps and housing vouchers are less efficient forms of redistribution than modifications to the income tax schedule. Put differently, the beneficiaries of these policies would prefer the government instead spend the same amount of money on a tax cut (e.g. EITC expansion). In contrast, the estimates suggest the JTPA may be a more efficient policy than a tax cut. As shown in Hendren (2016), spending on the JTPA generated large fiscal externalities through increased taxable income and reductions in other social programs, which effectively lowers the net cost of the policy to the government, c.[29]

Figure 8 provides a graphical method for assessing the efficiency of policies by comparing their MVPF to the MVPF of a tax cut with a similar distributional incidence. In particular, it is not efficient to ask whether low-income populations are willing to pay the cost of the policy. Rather, it is efficient to ask whether they prefer additional spending on the policy relative to spending the same amount on a tax cut targeted to the same individuals. Because it is costly to provide transfers to the poor and cheap to provide them to the rich, the results suggest one should use a threshold for the MVPF that is below 1 for low-income beneficiaries and above 1 for high income beneficiaries.[30] More generally, one can use the framework here to test for the efficiency of government policies in a manner that does not rely on a social welfare function. Normative assessments of the desirability of economic policies are turned into positive assessments of causal effects and individuals' willingnesses to pay.

8 Conclusion

In their original work, Kaldor and Hicks hoped to provide a method to avoid the inherent subjectivity involved in resolving interpersonal comparisons. Weighting surplus using efficient welfare weights measures the economic efficiency of an alternative environment (or policy change) using the Kaldor-Hicks redistributive experiments but accounting for the distortionary cost of taxation. Estimates for the US suggest that redistribution from rich to poor is more costly than from poor to rich. Thus, it is efficient to place greater weight on the poor than on the rich. Regardless of one's own social preferences, surplus to the poor can be turned into greater welfare for everyone than surplus to the rich. The shape of the efficient welfare weights is largely driven by the shape of the income distribution, as opposed

[28] Incomes are deflated to 2012 dollars using the CPI-U.

[29] Hendren (2016) discusses the conceptual issues in measuring the willingness to pay for the JTPA program. Figure 8 presents the baseline estimates of the MVPF for the JTPA that assumes individuals value the program at its program cost. Hendren (2016) discusses several alternative methods for recovering this willingness to pay; on the one hand, beneficiaries saw large increases in earnings, which suggests they may have substantial willingness to pay for the program above its program cost. This would be the standard approach to measuring the benefits in previous literature (e.g. Bloom et al. (1997)). However, this assumes no disutility of effort changes for those who increased their earnings. Moreover, enrollees were allowed to enter for free, which implies one can not in principle rule out arbitrarily small willingnesses to pay for the program.

[30] Moreover, recall that the conclusion about whether $g(y)$ is above or below one does not depend on the exact specification of $\epsilon^c(y)$, but rather depends largely on the shape of the income distribution, $\alpha(y)$. In this sense, the conclusion is robust to a wide range of assumptions about the behavioral impact of taxes.

to assumptions about behavioral responses to taxation. As a result, the broad conclusion of declining efficient welfare weights is robust to a wide range of assumptions about behavioral elasticities.

There are many important directions for future work, including incorporating the general equilibrium effects of taxation (as in ongoing work by Tsyvinski and Werquin (2018)). Additionally, one could extend the analysis here to construct weights that involve redistribution not just through the tax schedule but also via other means, such as health insurance subsidies or other policies. By expanding the dimensionality of the weights, it could help deal with settings where surplus varies conditional on income.

In the end, reasonable people and economists will always disagree about the optimal degree of redistribution in society. But, such debates need not lead to paralysis for debates about how best to bring about this degree of redistribution. To that aim, I hope the efficient welfare weights can be a tool to help generate greater consensus about the desirability of economic policies.

References

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. 7.2, 7.2, 29

Bourguignon, F. and A. Spadaro (2012). Tax-benefit revealed social preferences. *Journal of Economic Inequality 10*, 75–108. 1, 3.3.1, 3.4

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2015). The effect of wealth on individual and household labor supply: evidence from swedish lotteries. Technical report, National Bureau of Economic Research. 5.3, 6.1, 24, E

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018. 6.1

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)*. 6.1

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity: The geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics 129*(4), 1553–1623. 5.1

Coate, S. (2000). An efficiency approach to the evaluation of policy changes. *Economic Journal 110*(463), 437–455. 1, 4, 3.1, 3.2

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. 5.3, D

Gruber, J. and E. Saez (2002, April). The elasticity of taxable income: evidence and implications. *Journal of Public Economics 84*(1), 1–32. 5.3, 6.1

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy 30*. 1, 23, 26, 7.2, 7.2, 29

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

Hicks, J. R. (1940). The valuation of social income. *Economica 7*(26), 105–124. 1, 2.1, 3.1, 3.2

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. 23

Hoynes, H. W. and D. W. Schanzenbach (2012). Work incentives and the food stamp program. *Journal of Public Economics 96*(1-2), 151–162. 7.2, 7.2

Hylland, A. and R. Zeckhauser (1979). Distributional objectives should affect taxes but not program choice or design. *The Scandinavian Journal of Economics 81*(2), 264–284. 1, 1, 7, 7.1, 7.1, G

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. 7.2, 7.2

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal 49*(195), 549–552. 1, 1, 2.1, 3.1, 3.2, F

Kaplow, L. (1996). The optimal supply of public goods and the distortionary cost of taxation. *National Tax Journal 49*, 513–534. 1, 1, 5, 7, 7.1, 7.1, G

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, 1, 5, 7, 7.1, 7.1, G

Kaplow, L. (2008). *The Theory of Taxation and Public Economics*. Princeton University Press. 1, 1, 5, 7, 7.1, 7.1, G

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. 16, 26

Liebman, J. and E. Saez (2006). Earnings responses to increases in payroll taxes. *Working Paper*. 6.1

Lockwood, B. B. and M. Weinzierl (2016). The evolution of revealed social preferences in the united states and the costs of unequal growth and recessions. *Journal of Monetary Economics 77*, 30–47. 7

Mayshar, J. (1990). On measures of excess burden and their applications. *Journal of Public Economics 43*(3), 263–89. 26

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

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

Piketty, T. and E. Saez (2013). Optimal labor income taxation. *Handbook of Public Economics 5*, 391. 5.3, D

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*, 205–229. 1, 2, 17, 32, D

Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*(3), 1039–1073. 1, 2

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. 5, 5.1, 5.3, 6.1

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45. 3, B

Schlee, E. E. (2013). Radner's cost-benefit analysis in the small: An equivalence result. *Working Paper*. 11

Slemrod, J. and S. Yitzhaki (1996). The social cost of taxation and the marginal cost of funds. *International Monetary Fund Staff Papers 43*(1), 172–98. 26

Tsyvinski, A. and N. Werquin (2018). Generalized compensation principle. Technical report, National Bureau of Economic Research, Working Paper 23509. 3.4, 8

Werning, I. (2007). Pareto efficient income taxation. *MIT Working Paper*. 6, 3.3.2, 22, 6.2

Whitmore, D. (2002). What are food stamps worth? *Princeton University Industrial Relations Section Working Paper #468*. 7.2

Zoutman, F. T., B. Jacobs, and E. L. W. Jongen (2013). Optimal redistributive taxes and redistributive preferences in the netherlands. *Mimeo*. 1, 2.2, 3.3.1

Zoutman, F. T., B. Jacobs, and E. L. W. Jongen (2017). Revealed social preferences of dutch political parties. *Journal of Public Economics 165*. 1, 3.3.1, 17

Appendix (Not for Publication)

A Proofs

A.1 Preliminaries

For all proofs below, let $\mu(\theta)$ denote the measure over the type distribution and let $F(y)$ denote the cumulative distribution of income in the status quo, $F(x) = \int 1\{y(\theta) \leq x\} d\mu(\theta)$. The function $\hat{T}(y; y^*, \epsilon, \eta)$ is given 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 let $\hat{q}(y^*, \epsilon, \eta)$ denote the net government resources expended under $\hat{T}(y; y^*, \epsilon, \eta)$. 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:

$$\hat{q}(y^*, \epsilon, \eta) = \frac{-1}{F\left(y^* + \frac{\epsilon}{2}\right) - F\left(y^* - \frac{\epsilon}{2}\right)} \int_\theta \hat{T}\left(\hat{y}(\theta; y^*, \epsilon, \eta); y^*, \epsilon, \eta\right) d\mu(\theta) \tag{9}$$

One needs to evaluate this derivative with respect to η at $\eta = 0$. WLOG, I assume $\hat{q}(y^*, \epsilon, 0) = 0$ so that the status quo tax schedule is budget neutral.

A.2 Proof of Proposition 1

Statement of Proposition *For any $\epsilon > 0$ define the scaled surplus by $s_\epsilon(y) = \epsilon s(y)$ and $S_\epsilon = E[s_\epsilon(y) g(y)] = \epsilon S$. If $S < 0$, 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, $s^t_\epsilon(y)$, that is uniformly greater than the surplus offered by the alternative environment, $s^t_\epsilon(y) > s_\epsilon(y)$ for all y. Conversely, if $S > 0$, no such $\tilde{\epsilon}$ exists.*

Proof The strategy of the proof is to consider a modification to the tax schedule that gives a discrete tax cut to each interval of the income distribution that makes everyone in the interval better off relative to the alternative environment. I show that when $S < 0$ and for sufficiently small ϵ, one can find sufficiently fine partitions that lead to feasible modifications of the tax schedule that make everyone better off relative to the alternative environment.

More formally, suppose $S < 0$. Then,

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

so that

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

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

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

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:

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

and let

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

Note that aggregate government revenue is the sum of revenue from the partition. Assumption 1, combined with the linearity properties of the derivative implies that

$$\frac{d}{d\epsilon}|_{\epsilon=0} R\left(\hat{T}_\epsilon^P\right) = \sum_{j=1}^{N_P} \frac{d}{d\epsilon}|_{\epsilon=0} R\left(\hat{T}_\epsilon^j\right) \tag{10}$$

Note that each partition can be represented as

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

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 $\Pr\left\{y\left(\theta\right) \in P_j\right\} = \mu\left\{y^{-1}\left(P_j\right)\right\}$ and $\hat{q}$ is defined in equation (9) above.

Now, define η_j^P as

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

where $\bar{g} = E\left[g\left(y\right)\right]$ is the average value of the marginal cost of taxation. Let $s_\epsilon^t\left(y\right)$ denote the surplus the individual earning y 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 $\epsilon\left(\sup\left\{y | y \in P_j\right\}\right)$

$$s_\epsilon^t\left(y\right) > \epsilon\left(\sup\left\{y | y \in P_j\right\}\right) \tag{11}$$

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

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

Note that the terms inside the sum have limits that exist and are unique (because $g\left(y\right)$ 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 $g\left(y\right)$ – some suitable integrability condition would work – but this is sufficient. So,

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

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 $s_{\epsilon}^{t}\left(y\right) > s_{\epsilon}\left(y\right)$ for all y.

Converse Now suppose $S > 0$. Then,

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

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, $\frac{d}{d\epsilon}|_{\epsilon=0} s_{\epsilon}^{t}\left(y\right) \geq s\left(y\right)$. I will show that this implies the tax schedule modification is not budget neutral for sufficiently small ϵ.

Note that the envelope theorem implies $\frac{d}{d\epsilon}|_{\epsilon=0} \hat{T}_{\epsilon}\left(y\right) = \frac{d}{d\epsilon}|_{\epsilon=0} s_{\epsilon}^{t}\left(y\right)$ for all 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 $E\left[s_{\epsilon}^{t}\left(y\left(\theta\right)\right) | y\left(\theta\right) \in P_j^{\gamma}\right]$ 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

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

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

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

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

Therefore,

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

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

$$\frac{d}{d\epsilon}|_{\epsilon=0}R\left(\hat{T}_{\epsilon}\right) \leq -SE\left[g\left(\theta\right)\right] < 0$$

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 (11) 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 (10)).

A.3 Proof of Proposition 2

Statement of Proposition *Suppose Assumption 2 holds. For* $\epsilon > 0$, *let* $s_\epsilon = \epsilon s(y)$. *If* $S > 0$, *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* $s_\epsilon^t(y)$ *that is positive at all points along the income distribution,* $s_\epsilon^t(y) > 0$ *for all* y. *Conversely, if* $S < 0$, *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, which may differ from their choice of y in the alternative environment. To capture this, 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 2, this function is a bijection. Given the surplus function, $s(y)$, consider a modification to the income distribution that taxes away all but $\epsilon\frac{S}{2}$ 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

$$\hat{T}_\epsilon(y) = \tilde{T}_\epsilon(y) + \epsilon\left(s\left((y_\epsilon^\alpha)^{-1}(y)\right) - \frac{S}{2}\right)$$

Let $s_\epsilon^t(y)$ 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 $S > 0$, so that there is aggregate surplus to spread around). Hence, $s_\epsilon^t(y) > s_\epsilon(y) - E[s_\epsilon(y)]$ for sufficiently small ϵ; and taking the expectation conditional on $y(\theta) = y$ yields

$$E\left[s_\epsilon^t(y)\right] > 0 \quad \forall y$$

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 $s_\epsilon\left((y_\epsilon^\alpha)^{-1}(y)\right)$. Note that Assumption 2, the marginal cost of implementing these surplus transfers is the same as in the status quo environment.

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

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

A.4 Proof of Proposition 3

This subsection provides the derivation of the marginal cost of taxation using elasticities when the utility function satisfies Assumption 3.

A.4.1 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 that allow types θ to move to/from 0 and a point of interior earnings, $y > 0$, in response to a change in the size of the tax cut, η.

A key source of complexity is that 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:

$$c'(y(\theta); \theta) = -\frac{u_y}{u_c} = 1 - T'(y(\theta)) \tag{12}$$

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^*,

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

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

Clearly,

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

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 $k(\theta)$. Formally, recall that $\mu(\theta)$ is the measure on the type space. Let $\mu_{\theta|k}(\theta|k)$ denote the measure of θ conditional on having curvature k (i.e. $c''(y(\theta)) = k$)

and let $\mu_k(k)$ denote the measure of those having curvature k.[31] Then,

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

Taking a derivative yields

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

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

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

Therefore, evaluating at $\eta = 0$ yields

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

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

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

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

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

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

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

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

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

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

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,

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

so that

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

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

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

A.4.2 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 (12)) held. Now, I show how one can overlay participation margin responses for people who 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

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

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

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

so that

$$\begin{aligned}
\lim_{\epsilon\to0}\frac{1}{\epsilon}\frac{dP}{d\eta}|_{\eta=0} &= E\left[\left(T\left(y\right)-T\left(0\right)\right)\frac{dLFP\left(y\right)}{dc}|y^{P}\left(\theta\right)=y\right]LFP\left(y\right) \\
&= \frac{T\left(y\right)-T\left(0\right)}{y-T\left(y\right)}\hat{\epsilon}\left(y^{P}\right) \\
&= \frac{T\left(y\right)-T\left(0\right)}{y-T\left(y\right)}\epsilon_{c}^{LFP}\left(y\right)
\end{aligned}$$

where $\hat{\epsilon}_{c}^{LFP}\left(y\right)$ is the semi-elasticity of labor force participation at y off of the base of all potential people who have $y^{P}\left(\theta\right)$ 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 $LFP\left(y\right)$, this is equal to the elasticity of labor force participation at $y^{P}\left(\theta\right)$

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

Therefore, we have

$$\lim_{\epsilon\to0}\frac{d\hat{q}\left(y^{*},\epsilon,\eta\right)}{d\eta}|_{\eta=0}=1+\frac{1}{f\left(y^{*}\right)}\frac{d}{dy}|_{y=y^{*}}\left[\frac{T'\left(y\right)}{1-T'\left(y\right)}\epsilon^{c}\left(y\right)yf\left(y\right)\right]-\frac{T'\left(y\right)}{1-T'\left(y\right)}\zeta\left(y\right)+\frac{T\left(y\right)-T\left(0\right)}{y-T\left(y\right)}\epsilon_{c}^{LFP}\left(y\right) \tag{13}$$

where

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

and

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

B Inverse Optimum Derivation

Efficient social welfare weights correspond to the implicit welfare weights 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, so that the marginal impact on social welfare of providing an additional \$1 of resources to type θ is $\chi(\theta)$, which is normalized so that $E[\chi(\theta)] = 1$. 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 (2016)).

Proposition 4. *Suppose the income tax schedule in the status quo, $T(y)$, maximizes social welfare and let $\chi(\theta)$ denote the local social marginal utilities of income. Then, the efficient welfare weights $g(y)$, 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 $g(y) = E[\chi(\theta) | y(\theta) = y]$. $\square$

C Heterogeneity

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. Fortunately, with a slight modification of the surplus function, one can use the efficient welfare weights 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 weighting not actual surplus, $s(\theta)$, but rather these min and max surplus functions conditional on income. In particular, let

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

and

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

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

Proposition 5. *Suppose $\overline{S} < 0$. 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 $\overline{S} > 0$, 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 1. □

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

Conversely, using Assumption 2, 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 6. *Suppose Assumption 2 holds. Suppose $\underline{S} > 0$. 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, $s_\epsilon^t(\theta) > 0$ for all θ.*

Proof. The proof follows immediately by providing surplus $\underline{s}_\epsilon(y) = \inf\{s_\epsilon(\theta) | y(\theta) = y\}$ instead of $s_\epsilon(y)$ in the proof of Proposition 2. □

In general, it can be the case that $\overline{S} > 0 > \underline{S}$, so that the potential Pareto criterion cannot lead to a sharp comparison between the status quo and the alternative environment.

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

Dealing with Heterogeneity in Practice When surplus is heterogeneous conditional on income, it may be the case that $\overline{S} > 0 > \underline{S}$. 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 $\underline{S} > 0$. Of course, this might be overly conservative. Second, one can use average surplus, $S = E[s(\theta) g(y(\theta))]$, 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 does not rely on the Pareto principle, but may be a useful application 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 $\overline{S} > 0 > \underline{S}$, 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. Efficient surplus is well-suited to addressing comparisons where the key source of heterogeneity is income.

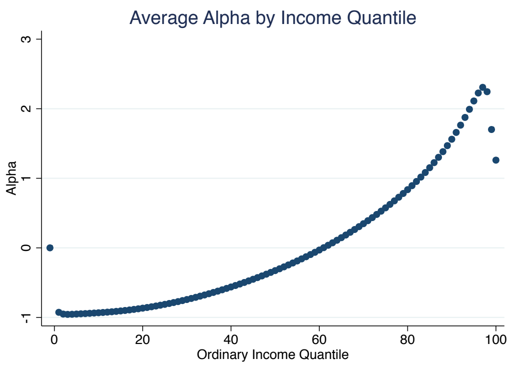
Appendix Table I
Summary Statistics

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

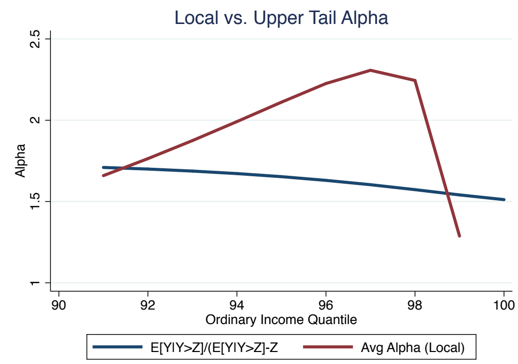
Notes. This table presents summary statistics for the universe of income tax returns in 2012 for U.S. citizens aged 25-60. This sample is used to construct the elasticity of the density of the income distribution (i.e. alpha) for each marginal tax rate. The table presents the number of filers, mean ordinary 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.

Appendix Figure 1: Estimation of Shape of Income Distribution

A. Average Value of $\alpha(y)$ by Income Quantile



B. Upper Tail Estimate of $\alpha(y)$, $\frac{E[Y|Y\geq y]-y}{y}$



Notes: Panel A of this figure presents estimates of the average value of $\alpha(y)$ by income quantile. Panel B presents estimates using an alternative method of estimating $\frac{E[Y|Y\geq y]-y}{y}$ in each quantile.

D Summary Statistics and Estimation Details

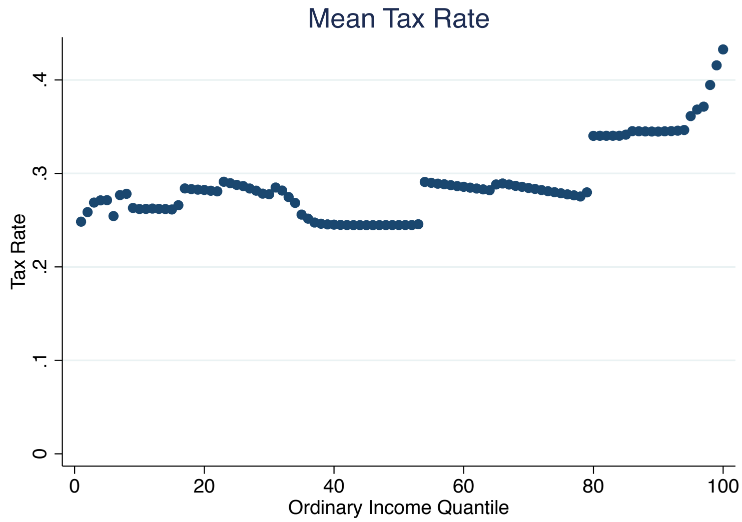
Appendix Table I presents the summary statistics of the sample used to construct the estimates of the shape of the income distribution conditional on the marginal income tax rate. Overall, there are roughly 100M filers aged 25-60 used in the analysis, with mean family incomes of roughly \$65M, and mean ordinary incomes of \$46M.

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 the shape of the income distribution, α, in a manner that allows it to vary with the marginal tax rate for a majority of the population. Let j index the set of tax schedules. 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, which I denote $\alpha_j(y)$.[32] 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.

[32] 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.

Appendix Figure 2: Average Tax Rates by Income Quantile



Notes: This figure presents the average tax rate by income quantile. Each tax rate is the sum of the federal income tax rate, state taxes, Medicare, sales taxes, and EITC top-up, as discussed in Section 5.1.

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.[33] 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. This allows for 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 the presence of multiple tax schedules.[34]

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

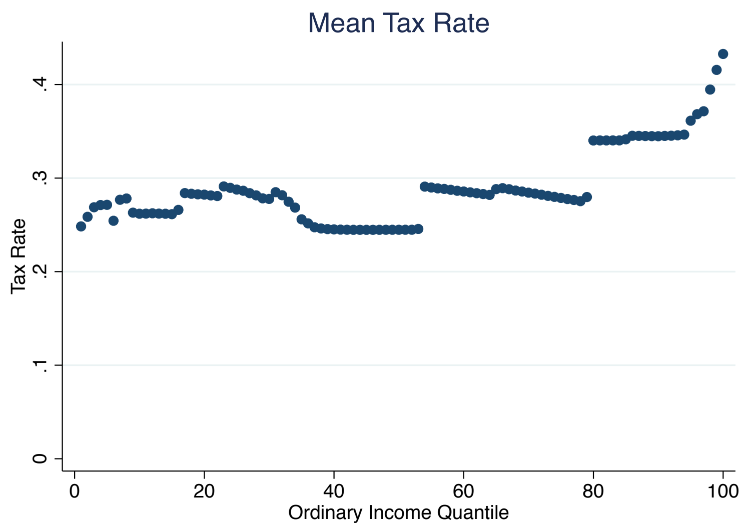
[33] This 500,000 threshold is chosen for computational simplicity on the remaining groups, but the results are similar to lowering it to 250,000.

[34] In practice, this degree of generality turns out not to matter in the estimation: one could arrive at a similar set of weights using the average Pareto parameter at each income level instead of estimating its heterogeneity across tax schedules.

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. Appendix Figure 1 (left panel) plots the average local estimate of α (using the fifth order polynomial) across the income distribution and Appendix Figure 1 (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 (2013). 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.

Appendix Figure 3: Incorporating Income Effects



Notes: This figure presents the efficient social welfare weights using both the baseline specification (solid blue line) and a modified specification that incorporates an income effect (dashed red line). To calculate the modified specification with the income effect, I assume a constant elasticity of labor supply with respect to income of -0.15, similar to the estimate in Cesarini et al. (2015).

E Income Effects

The baseline specification assumes no income effects on labor supply. This section illustrates how income effects increase the marginal cost of taxation, $g(y)$, but do so similarly at all points of the income distribution (assuming a constant elasticity). To illustrate, Appendix Figure 3 presents the baseline specification for $g(y)$ combined with an alternative specification that incorporates income effects. For simplicity, I approximate the income effect as $\zeta(y)\frac{\tau(y)}{1-\tau(y)}$ where $\tau(y)$ is the average marginal tax rate for those in each quantile of ordinary income. For $\zeta(y)$, I take an estimate of 0.15 from Cesarini et al. (2015) who study the impact of winning the lottery in Sweden on labor supply.

As shown in Appendix Figure 3, incorporating income effects raises the marginal cost of taxation at all income levels. But, in contrast to the substitution effect and the compensated elasticity, it does not differentially affect the marginal cost of taxation at different income levels. In this sense, the broad set of conclusions that one should apply greater weight to surplus tot he poor than to the rich remains true if one incorporates income effects into the analysis.

F 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)).

Kaldor and Hicks' original motivation was the comparison of different distributions of endowments. Motivated by this classic comparison, this Appendix uses the efficient welfare weights to compare distributions 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.

To compare income distributions, one needs to define a conceptual experiment that clarifies where an individual in one distribution would fall in the alternative distribution. This experiment then defines the surplus function, $s(y)$, that can be used to compare the distributions.

In general, one could consider any number of potential experiments. Perhaps 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 $[0,1]$. This generates the surplus function:

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

For simplicity, 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. This minimizes the size of each individual surplus, which helps make the first-order approximation for the marginal cost of taxation more appropriate. Moreover, choosing $r(\alpha) = \alpha$ minimizes the estimated surplus of the status quo relative to the alternative environment. 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.

F.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 1 plots several quantiles of the household after-tax income distribution over time using data from the Congressional Budget Office (CBO) from 1979-2009.[35] 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.

[35] 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

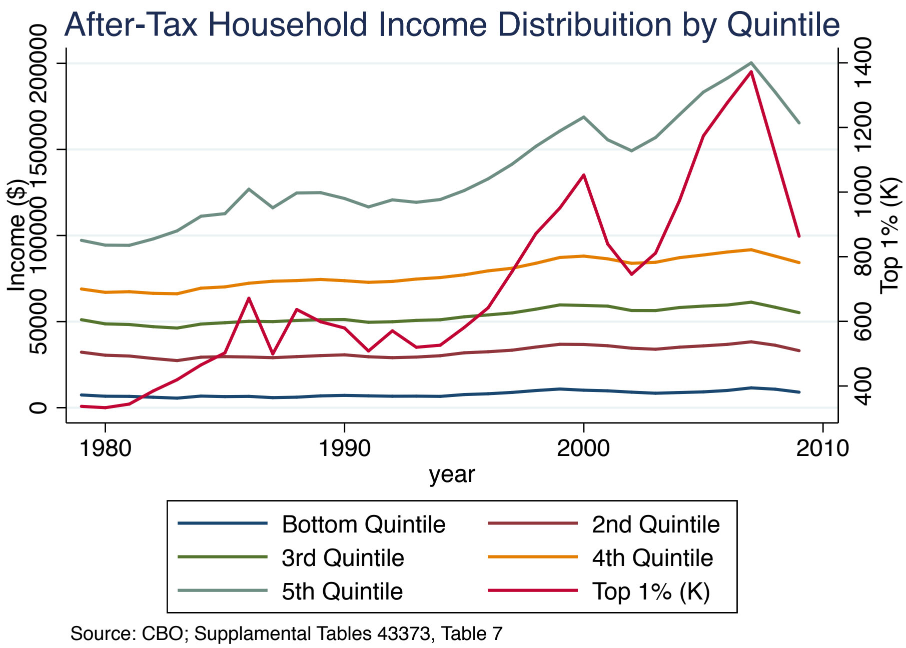


Figure 1: Figure presents the evolution of the distribution of after-tax income in the U.S. from 1979-2009 using CBO data.

Here, I use the efficient welfare weights 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 efficient surplus in household income by

$$S_t = \int_0^1 \left[Q_0(\alpha) - Q_t(\alpha)\right] g^H\left(Q_0(\alpha)\right) d\alpha \tag{14}$$

where $g^H(y)$ are the efficient welfare weights. Intuitively, S_t 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. All incomes are in units of 2012 income using the CPI-U deflator.

Figure 2 reports the change in mean household income (dashed blue line), along with the efficient surplus under the baseline specification and two alternative elasticity specifications. Mean household income has increased by roughly \$18,300 relative to 1980, but if these benefits were redistributed equally across the population, growth would have 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,

of income.

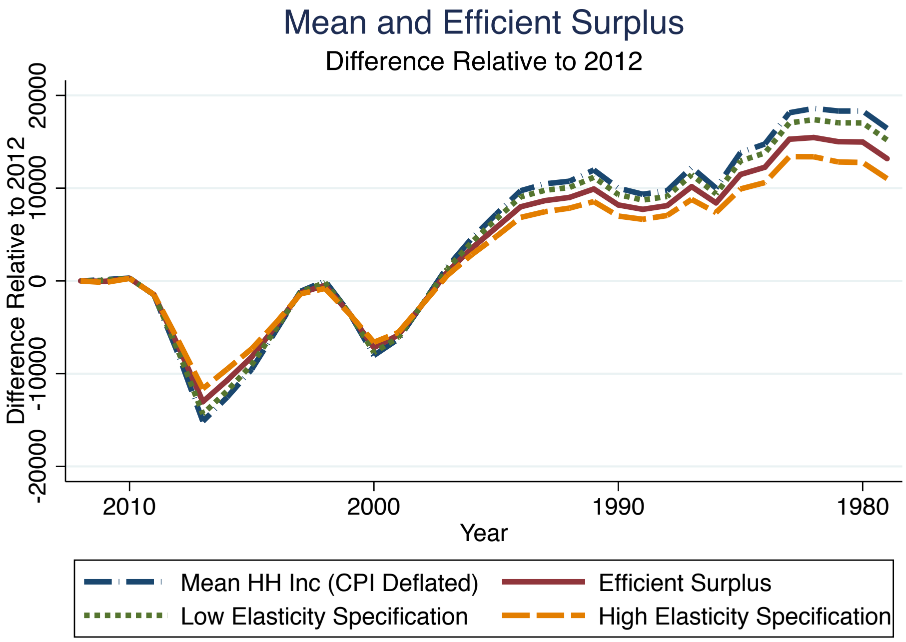


Figure 2: Raw and Deflated Household Income Change Relative to 2009 (All Income deflated to 2009 dollars using the CPI)

roughly 15-20% of the growth since 1980 would be evaporated.

Figure 3 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.[36] 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. 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.

F.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. In this subsection, I use the efficient welfare weights to ask how much richer or poorer the U.S. would be relative to these countries if it attempted to replicate their income distributions using modifications to the tax schedule.

The efficient surplus associated with moving from the status quo income distribution to the income

[36]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.

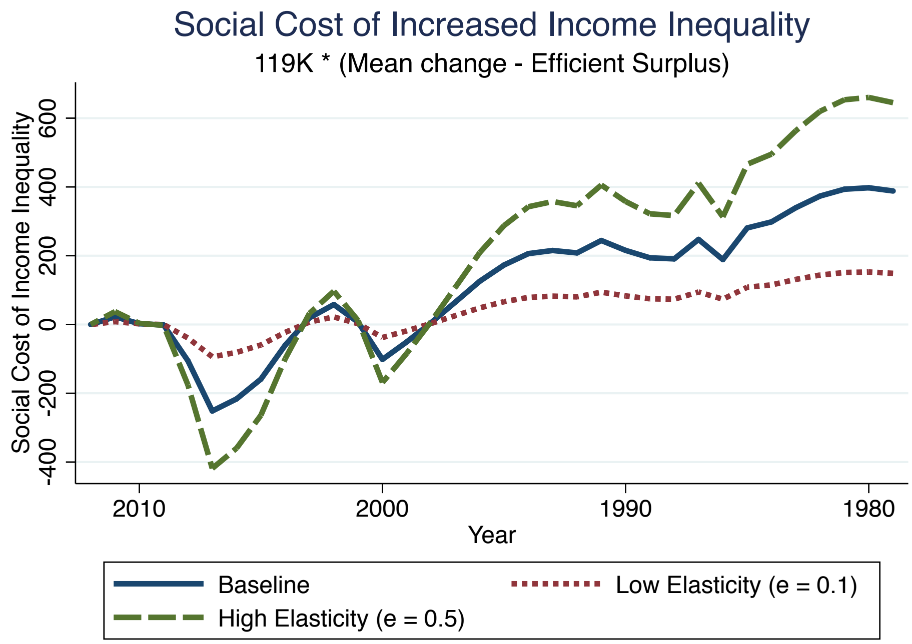


Figure 3: Social Cost of Increased Income Inequality

distribution in country a is given by

$$S_a^{ID} = \int_0^1 \left[Q_a\left(\alpha\right) - Q_0\left(\alpha\right)\right] g^H\left(Q_0\left(\alpha\right)\right) d\alpha \tag{15}$$

I form estimates of $Q_a\left(\alpha\right)$ using 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.

Figure 4 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 efficient welfare weights 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. But 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, it would be efficient to take Finland's income distribution over the US's income distribution, even though it has \$3,180 less in per capita national income.

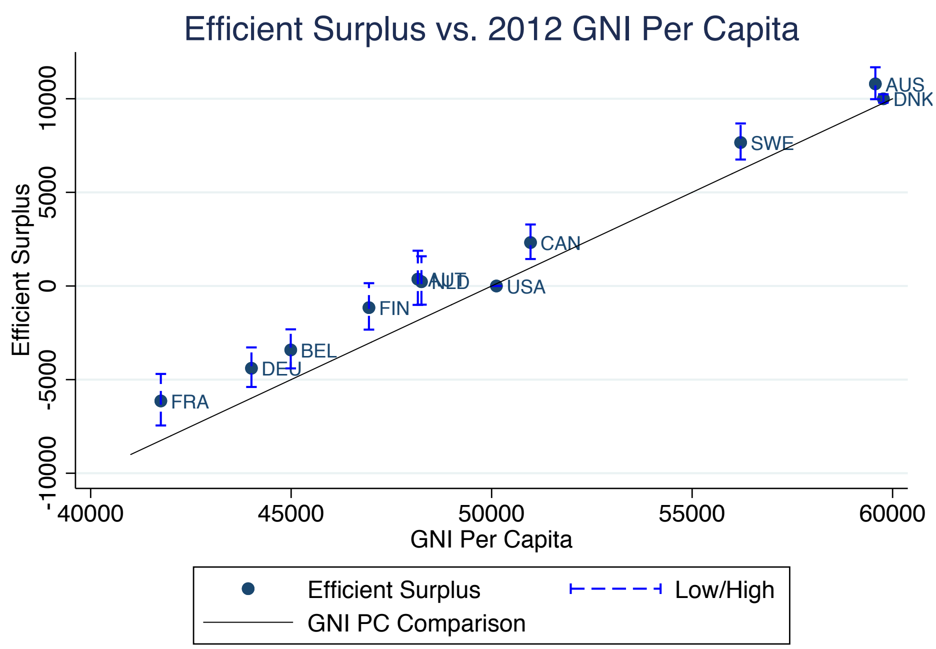


Figure 4: For each country, the efficient surplus (defined in equation 15) is presented for the baseline elasticity specification against GNI per capita on the horizontal axis; vertical bars representing the high and low elasticity specifications. If all countries had the same degree of inequality, then 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.

G Testing Weak Separability: Relation to Kaplow (2000, 2006, 2008, etc.) and Hylland and Zeckhauser (1979)

There is an long debate about whether or not one should weight the willingness to pay for publicly provided goods for the poor differently than the rich. Most influentially, Hylland and Zeckhauser (1979) followed by Kaplow (1996, 2004, 2008) provide a weak separability assumption on the utility function that, if satisfied, implies that additional spending on the publicly provided good increases utility if and only if the sum of individuals' willingness to pay exceeds the mechanical cost of the publicly provided good. This Appendix shows how this theoretical result is nested in the model of Section 7, and thus the welfare framework provides a test of the weak separability assumptions employed in this literature.

Consider a policy of spending \$1 per capita on a publicly provided good, G. This will have a net cost to the government of $\$c$ that may differ from \$1 because of any fiscal externalities from behavioral responses to the provision of the public good, $FE^G = c - 1$. Assume that individuals of income level y are willing to pay $s(y)$ for this additional expenditure so that the average willingness to pay is $E[s(y)]$. Individuals are thus willing to pay the mechanical cost of the expenditure if and only if $E[s(y)] \geq 1$. In contrast, equation (6) (generalized to the case of willingness to pay that varies with y) suggests that additional spending on G is efficient if and only if $E[g(y)\,s(y)] \geq c$. How are these different?

It turns out that the weak separability assumption in Hylland and Zeckhauser (1979) and Kaplow (1996, 2004, 2008) implies that the behavioral response to \$1 of a tax cut to those earning near $\$y$ should be the same as the behavioral response to a policy that provides \$1's worth of additional G to those earning near $\$y$. Weak separability imposes that the behavioral response to a tax cut scaled by the willingness to pay for the policy equals the behavioral response to the policy. Hence,

$$FE^G = E[s(y)\,FE(y)] \tag{16}$$

where $s(y)$ is the willingness to pay of individuals with income y for the additional spending on G and $FE(y)$ is the fiscal externality associated with the tax cut. Hence, the total cost of the additional spending on G is equal to

$$\begin{aligned} c &= 1 + FE^G \\ &= 1 + E[s(y)\,FE(y)] \end{aligned}$$

Hence, testing whether $E[s(y)\,g(y)] \geq c$ is equivalent to testing whether

$$E[s(y)\,g(y)] \geq c \iff E[s(y)\,(1 + FE(y))] \geq 1 + E[s(y)\,FE(y)] \iff E[s(y)] \geq 1$$

So, the test for efficiency reduces to

$$E[s(y)] \geq^{?} 1$$

which asks whether the aggregate willingness to pay exceeds the mechanical cost of the policy (that does not include the fiscal externalities). In this sense, if equation (16) holds for the policy change in question, one need not know either the efficient welfare weights, $g(y)$, or the fiscal externalities induced by the policy change, FE^G. One can simply compare unweighted aggregate willingness to pay to the mechanical cost of the policy. But more generally, testing whether $E[g(y)s(y)] \geq c$ provides a general method for asking whether the policy is efficient, even if weak separability does not hold.