Abstract

There are increasing calls for concrete suggestions for how to account for distributional impacts in regulatory analysis. Within the context of benefit-cost analysis, per se, one possibility is to apply “distributional weights,” to inflate costs and benefits experienced by poor or disadvantaged groups. We distinguish between “welfare weights,” intended to correct for the bias in willingness to pay caused by diminishing marginal utility of income, and “equity weights,” intended to account for the possibility that decision makers might place greater moral weight on the welfare of the poor, or other disadvantaged groups. We argue that welfare weights are appropriate and necessary to maintain the legitimacy of BCA as a measure of aggregate welfare, but that equity weights are inappropriate, because they involve moral judgments that should remain in the domain of democratically accountable decision makers, and obfuscate both the welfare and equity impacts of policies. We offer concrete suggestions regarding the application of welfare weights, and the calculation of “breakeven equity weights,” to provide intuitively comprehensible and useful information about the tradeoffs between welfare and equity implicit in policies.

Section 1: Introduction

The importance of equity and distributional impacts in rule-making within the federal government of the United States has increased over time. Executive Order 12291, signed by Ronald Reagan in 1981, established the requirement that major rules be subjected to regulatory impact analysis. No mention was made of either equity or distributional concerns. In 1993, when Bill Clinton signed Executive Order 12866, updating Reagan’s order, equity and distribution were mentioned explicitly as pros and cons of rules and regulations that agencies were required to consider. Joe Biden has pushed these concerns even more

---

1 Loomis (2011) and Nurmi and Ahtiainen (2018) review the increasing number of places where analysis of distributional consequences is required.
2 Circular A-4, the document that guides executive agencies in complying with the Executive Order, states this requirement explicitly: “Your regulatory analysis should provide a separate description of distributional effects (i.e., how both benefits and costs are distributed among sub-populations of particular concern) so that decision makers can properly consider them along with the effects on economic efficiency.”
into the forefront in a memorandum released on the first day of his presidency, titled “Modernizing Regulatory Review.” The memorandum directs the Director of the Office of Management and Budget to “propose procedures that take into account [among other things] the distributional consequences of regulations, including as part of any quantitative or qualitative analysis of the costs and benefits of regulations,” and to “provide concrete suggestions on how the regulatory review process can promote public health and safety, economic growth, social welfare, racial justice, environmental stewardship, human dignity, equity, and the interests of future generations.” This goes considerably beyond previous guidance, as it not only requires that agencies consider equity and distributional concerns, but that they be required to use consistent and explicitly defined quantitative methodologies for doing so. The purpose of this paper is to provide the kind of concrete suggestions we believe the memorandum is calling for.

Regulatory review does not need to be strictly limited to Benefit-Cost Analysis (BCA), but the purpose of this paper is to propose a set of procedures for accounting for distributional consequences within BCA per se. One approach that has been proposed is to introduce “distributional weights” into BCA, by multiplying the monetized value of impacts on the poor by some number greater than one, and those on the wealthy by some number equal to or less than one. The idea of distributional weighting of impacts at different levels of income has been discussed by scholars for many years, with the pace of publication accelerating in the 21st Century, and particularly the last decade. Distributional weights have been proposed as a way to address two separate issue, and one of the most important points we hope to make is that these two issues must be understood, and treated, differently. The first is that Willingness to Pay (WTP), as a measure of welfare, is biased against the poor, because a dollar of consumption generates more welfare

---

4 Our discussion will be largely restricted to distribution of impacts across income groups, but our approach can be applied to distribution across racial groups, or other distributional categorizations.
for a poor person than a wealthy person, so that any given change in the welfare of the poor will be represented by a smaller number of dollars in BCA than the same change in the welfare of the wealthy. Thus, BCA actually promotes inequality, the opposite of what the Biden memorandum calls for. If we knew how much more a dollar matters to a poor person than to a wealthy person in quantitative terms, we could adjust the monetary value of welfare impacts so that a given change in welfare is represented by the same number of dollars, regardless of income.

The second reason one might apply weights to BCA by income is that decision-makers may simply “care” more about how policies affect the welfare of the poor than that of the wealthy. A dramatically unequal distribution of welfare is considered by many to be unjust, unfair, or simply disgusting. Even if WTP were adjusted to overcome the bias referenced above, BCA would still place equal weight on the WTP of all groups, regardless of income. As such, the aggregate net benefit of a policy measured by BCA, even after adjusting for diminishing marginal utility of income, tells decision-makers nothing about the distributional impacts of the policy. In the parlance of the Biden memorandum, there is no quantitative accounting within BCA of the distributional pros and cons of the policy, and thus no way for decision-makers to distinguish policies with beneficial distributional impacts from those with deleterious distributional impacts. Overweighting the welfare of the poor and under-weighting that of the wealthy is seen as a way to overcome this problem, turning BCA into a tool that can, in the parlance of the Biden memorandum, “promote equity.”

The thesis of this paper is that these two reasons for applying weights to costs and benefits in BCA must be understood as separate, and consequently, they should be treated very differently. Correcting for the bias in WTP as a measure of welfare is not a way of accounting for the distributional impacts of a policy. It is simply a way of keeping BCA from being positively biased against the poor. It is fundamentally a
technical concern in the measurement of welfare and, as such, does nothing to account for distributional impacts per se. It is only when we additionally overweight the welfare of the poor for reasons of conscience and moral sensibility that we are truly accounting for distributional impacts. One of the important distinctions between these two issues is that the former, being a technical matter, belongs under the purview of the analyst, while the latter, being a matter of moral value, requires a moral value judgment, which puts it under the purview of democratically accountable decision-makers. Furthermore, simply correcting for bias keeps BCA entirely within the domain of welfare, so that it is only measuring one value, and provides clear and unambiguous information about that value, while weighting to account for society's particular concern for the welfare of the poor draws BCA into the moral domain of equity, a quite different kind of value. Attempting to account for welfare and equity with the same, unitary cardinal measure obfuscates the impacts of a policy on these two domains. As we will show, once BCA has been weighted to address equity concerns, it is no longer possible to extract any information from it regarding either the welfare impacts or the equity impacts of a policy.

For these reasons, we recommend that adjusting for bias should be done, and should be done by analysts, while weighting for true distributional concerns should not be done at all. Instead, we propose an extremely simple method for extracting useful quantitative information about the distributional impacts of a policy from BCA, which is to compute the breakeven weight on the costs and benefits of the poor that would cause a policy to switch from being net negative to being net positive in terms of net benefit, or vice versa. In other words, if a policy generates costs(benefits) for the poor but has positive(negative) net benefits, how much more would we have to care about the poor, relative to the wealthy, in proportional terms, for equity concerns to outweigh welfare concerns.
In section 2 of the paper, we further explicate the distinction between bias caused by diminishing marginal utility of income and true equity concerns. In section 3, we make concrete recommendations for the use of weights in BCA. Section 4 concludes.

Section 2: Differentiating between welfare weights and equity weights.

2.1 Welfare weights.

In the literature on distributional weighting in BCA, the distinction between the issue of WTP being a biased measure of welfare and the issue of BCA placing equal weights on the welfare of different groups, has been murky at times. In an early paper by Arnold Harberger (1978) titled "On the use of distributional weights in social BCA," the discussion is entirely about correcting for diminishing marginal utility, with no recognition that this does not actually address concerns about the distribution of welfare. Adler (2008) spends five pages dissecting distributionally weighted BCA without noting the distinction, instead describing weighted BCA exclusively as a way to address equity concerns. Then, in a later paper entirely devoted to distributionally weighted BCA, Adler (2013) is very clear about the distinction between the two issues, but ultimately settles on a set of weights that addresses both issues at once, thereby rendering the distinction impossible for policy-makers to discern. Hammit (2020) writes about both equity considerations and the problem of bias caused by diminishing marginal utility of income without making any clear distinction between the two. Nurmi and Ahtiainen (2018) are very clear on the distinction but claim that it is better to address the two issues simultaneously, with a single set of weights, because, among other things, "the results are easier to interpret." We will demonstrate, later in the paper, that in fact, precisely the reverse is true. When the two issues are conflated by using a single set of weights, the interpretation is entirely opaque.
When the bias caused by the fact that a dollar is worth more to a poor person than a wealthy person is addressed in the literature, there is unanimous agreement on the solution, which is to derive weights for different individuals based on their marginal utility of income. (Cowell and Gardiner 1999, Adler 2013, Nurmi and Ahtiainen 2018) Such weights are typically referred to as “welfare weights.” The marginal utility of income is the number of “utils” an individual gets from a dollar of income, and thus converts dollars into utils. Because the marginal utility of income is greater for a poor person than for a wealthy person, the conversion results in an unbiased measure of utility.

However, if the weight for each individual is simply their marginal utility of income, the results are no longer interpretable, because utility is scale invariant, which means that the same marginal utility of income could be represented by different numbers, depending on the specific utility function used to represent preferences over income. To see this clearly, recall that in expected utility theory, the utility index is unique up to an affine transformation, which means, for example, that if one analyst used some function, $U(y)$, capture preferences over income, and derived marginal utility weights by taking the derivative of that function, another analyst could just as easily capture the same preferences using $2U(y)$, and wind up with a set of marginal utility weights twice as large as the first analyst.

The standard solution is to select some reference income level (median income being the most obvious choice) and compute weights by dividing the marginal utility at the median income by the marginal utility at other income levels. (In our example above, the 2 would cancel out, so that the two analysts would wind up with the same weights.) This converts the dollar value of any given amount of welfare to a poor person into the dollar value of the same amount of welfare to a median-income person. The results are

---

5 The term welfare weights has also been used to refer to weights that additionally address equity concerns (Fleurbaey and Abi-Rafeh (2016), but it is more often used in the more limited way we are using it, as in Cowell and Gardiner (1999) and the U.K. Green Book (H.M. Treasury, 2018).
now straightforwardly interpretable. For any given income group, the welfare-weighted WTP represents the WTP they would express, for the welfare impact they experience, if they had the median income. Another interpretation is that the welfare-weighted net benefit to society of a policy is the net benefit that would be generated by implementing the policy in a hypothetical world where everyone had income equal to the median.

Inasmuch as BCA is often thought of as a measure of economic efficiency, it is important to distinguish between two concepts of efficiency, one reflected in unweighted BCA, the other reflected in welfare-weighted BCA. There is a fundamental conflict between two different interpretations or definitions of BCA. On the one hand, it is often referred to as a measure of whether a policy satisfies the potential Pareto criterion (e.g., Boardman et al., 2018), which is determined by taking the difference between the aggregate WTP of those who gain from a policy and the aggregate WTA of those who lose. To satisfy this definition of BCA, WTP must not be weighted, because if it is, it will, in general, no longer answer the potential Pareto question. On the other hand, BCA is often referred to as a measure of aggregate welfare (Harberger 1978, Adler and Posner 2006), in which case, WTP must be welfare-weighted, because WTP does not correctly measure welfare unless diminishing marginal utility is addressed. In light of these conflicting definitions of BCA, we propose a new nomenclature. We will refer to aggregate unweighted BCA as a measure of what we will call “potential Pareto efficiency,” and welfare-weighted BCA as a measure of what we will call “welfare efficiency. Using these terms drives home that unweighted BCA is implementing a particular moral criterion, the potential Pareto criterion, while welfare-weighted BCA is implementing a different moral criterion, maximization of aggregate welfare. The term welfare weights drives home the fact that these weights address a welfare issue, not an equity issue. Welfare weights

---

6 For example, consider a policy that generates a benefit to the poor of 10 and a cost to the wealthy of 15. If the welfare weight on the poor is 2, the policy will generate positive net benefit in terms of welfare but will fail the potential Pareto test.
simply create an unbiased measure of welfare. They do not in any way account for the differential welfare impact of a policy on different groups, or the fact that we might simply have greater concern for the welfare of certain groups.

A crucially important question we feel needs to be answered is, who should make the decision about what welfare weights to use? The heart of the matter, we feel, is the distinction between technical judgments and value judgments. How to weigh the welfare of the poor against the welfare of those better off—that is, how to make judgments about equity—is unambiguously a value judgment. It is not something that an economist can tell you, and thus, we feel, belongs unambiguously in the domain of democratically accountable decision-makers, whether appointed by the executive and approved by the legislature, or directly elected. In light of the Biden memorandum, it is clear that analysts must provide information about distributional impacts, of policies, but analysts should be given neither the responsibility, nor the authority, to make judgments about the relative importance to society of the welfare of the poor versus the welfare of the wealthy. However, we view the issue of diminishing marginal utility as a technical problem. Our goal as benefit-cost analysts is to base the quantitative measure of costs and benefits on the preferences of individuals. But preferences over income are not correctly measured by WTP. This is a technical flaw. As such, the question of what welfare weights to use is something that an analyst should, in principle, be able to determine. Alfred Marshall rendered his opinion in 1885, saying “taking account of the fact that the same sum of money measures a greater pleasure for the poor than for the rich [is a] task [that] most properly belongs to the economic organ.” (Marshall 1885) As we will discuss later, many economists have attempted to provide empirical evidence on diminishing marginal utility, using ostensibly objective methodology.
2.2 Equity weights.

Once BCA has been converted into an unbiased measure of welfare efficiency by the application of welfare weights, it estimates the *unweighted* sum of the welfare impacts on different groups. As such, it places the same moral weight on the welfare of individuals at different levels of income, and thus does not account for the possibility that decision makers, or society as a whole, may simply care more about the welfare of some groups than others. We believe that this is the issue that the authors of Circular A-4 had in mind when they carved out a special place within regulatory impact analysis, outside BCA, for “distributional concerns,” and what the authors of the Biden memorandum had in mind when they called for suggestions on ways that regulatory impact analysis could promote “equity.” This is not what Marshall (1885) or Harberger (1978) had in mind, but more recently, calls for weights that directly address equity concerns have entered the literature. Adler (2013) and Nurmi and Ahtiainen (2018) both call for using weights to address what the former referred to as diminishing “marginal moral value of utility.” This terminology immediately drives home the fact that this kind of weighting involves moral value judgments.

A considerable literature attempts to estimate the “correct” weights to use to address equity issues. Johansson-Stenman et al (2002) use experiments in which subjects are given choices between societies with different income distributions to estimate a parameter they call the coefficient of “social inequality aversion.” Amiel et al (1999) and Pirttilä and Uusitalo (2010), among others, estimate the same parameter, using a “leaky bucket” experiment, in which subjects are asked whether different hypothetical transfers of income from a wealthy person to poor person are acceptable.\(^7\)

---

\(^7\) There is a literature that uses political decisions to infer the inequality aversion of actual decision-makers. Perhaps the earliest attempt to do this is Eckstein (1961), which suggested using marginal income tax rates to determine weights for different income classes. For example, at the time the article was written, federal income tax rates for higher income families were double those for lower income families, implying, according to the article, that Congress believed taking a dollar from a low income family was equivalent to talking two dollars from a high income family. This approach seems to produce a mixed welfare and equity weighting scheme. Weisbrod (1968) develops an approach based on public expenditure decisions that perhaps comes closer to pure equity weights. To illustrate, say there are two projects, A and B. Suppose A has a larger net present value (NPV) but is rejected in favor of B because B provides greater net gains for the poor than A. Given these circumstances, and if the NPVs of each project is known,
Weights derived from this kind of data actually account for diminishing marginal utility of income and diminishing marginal moral value of welfare simultaneous, because they involve the distribution of income, not welfare. Adler (2013) describes a leaky bucket experiment involving transfers of welfare rather than money. Data from such an experiment would allow for identifying weights to be placed on the welfare of individuals at different income levels in order to account for equity. Thus, we would have a set of welfare weights to address marginal utility of income, and a set of what we refer to as “equity weights” to address diminishing marginal moral value of welfare. Clearly, if one were trying to use weights to address equity issues, one would want to simultaneously account for the bias in WTP as a measure of welfare, so that a single set of weights that addressed both issues would be appropriate. However, for our purposes it is crucial to think of welfare weighting and equity weighting separately, as we feel that the two reasons for weighting must be handled very differently.

Studies that attempt to identify “correct” equity weights are essentially attempts to measure the moral judgments of the citizenry, but a responsible decision-maker may or may not feel bound to assess policies on the basis of the moral judgments of the citizenry. Since J.S. Mill (1848), there has been a reasonable degree of consensus that, in a liberal society, the relative importance individuals place on the various ways weights for the poor and nonpoor ($W_p$ and $W_n$, respectively) can be derived by solving the following simultaneous equations, under the assumption that the decision-makers viewed the two projects of equal value:

$$\text{NPV}^A = W_p \cdot \text{NPV}_p^A + W_n \cdot \text{NPV}_n^A$$

$$\text{NPV}^B = W_p \cdot \text{NPV}_p^B + W_n \cdot \text{NPV}_n^B$$

These two approaches differ from the progressivity of taxes approach that, as discussed in Appendix 1, has been used to estimate the marginal utility of income, but all three are subject to the strong possibility that, in determining tax rates, decision-makers are concerned with factors other than distributional considerations, such as responding to lobbyist who represent the well-to-do. Yitzhaki (2003) develops weights based on the behavior of economists, rather than political decision-makers, by using inequality indices, such as Gini coefficients, commonly used in economic analysis. The paper argues on page 326 that “by choosing an inequality measure the economist implicitly assumes distributional weights.”
they allocate their own resources in pursue their own welfare (their WTP) ought to be given deference by decision-makers. In other words, when it comes to welfare, decision makers should not impose their own moral judgments. But when it comes to equity, we feel that decision makers ought not be constrained by the moral judgments of the citizenry but must make their own assessment of the moral impact of how society allocates welfare across individuals. As such, economists and analysts are not in a position to definitively determine the correct weights to use to address equity, and can, at best, provide insight into the attitudes of the citizenry. The decision of what weights to use—if any—must remain in the domain of democratically accountable decision-makers. Adler (2013) essentially makes this point.

“The use of distributional weights does raise questions of institutional role. An unelected bureaucrat might feel that it would be legally problematic, or democratically illegitimate, for her to specify weights. Who in government gets to act on contestable moral preferences is a complicated (and itself contestable) question of law and democratic theory. Suffice it to say that the advice welfare economists and moral theorists provide about the specification of weights is addressed to officials with the legal and democratic authority to act on such advice—whoever exactly those officials may be.”

**Section 3: Recommendations.**

In this section, we make the case for a concrete set of recommendations for how to address welfare weighting and equity weighting in BCA. We evaluate the pros and cons of different approaches using certain principles, or criteria that we consider important for policy analysis. We consider how to make use of welfare weights, what weights to use, and how to present and interpret unweighted and welfare-weighted net benefits. Separately, we explore the use of equity weights, and conclude that applying actual numerical equity weights is not appropriate, but computing breakeven equity weights can provide decision-makers with intuitively comprehensible and useful information for accounting for equity concerns.
3.1 What to do about welfare weights.

As mentioned earlier, there is a fundamental conflict between interpreting BCA in terms of aggregate welfare, which requires that WTP must be welfare-weighted, and interpreting it in terms of the potential Pareto criterion, which requires that WTP must not be welfare-weighted. A strong case has been made that, in the absence of actual transfers, the fact that the winners could compensate the losers and still come out ahead is a decision making criterion with little or no moral validity. The fundamental problem is precisely the diminishing marginal utility of money. Because wealthy people can pay more for their welfare than poor people (i.e., value the marginal dollar less), their welfare speaks louder than the welfare of the poor in the application of the potential Pareto criterion. As a consequence, a policy that benefits the wealthy and harms the poor could pass the potential Pareto criterion while actually making the poor worse off by more than it makes the wealthy better off. To see this most starkly, the potential Pareto criterion suggests that taking $100 from a wealthy person and giving it to a poor person is a "wash," and society as a whole is neither better or worse than the status quo. But, even if we do not consider the welfare of the poor to be more important to society than the welfare of the wealthy, it is intuitively obvious that the transfer has made the poor better off by more than it has made the wealthy worse off because the money matters more to the poor. These comments simply state the obvious. A metric that is biased against one group in favor of another is morally indefensible.

Two attempts at a defense have been mounted. The first is that if a policy satisfies the potential Pareto criterion, then policy makers could, if they chose, use the tax system to actually convey the necessary compensation from one group to another. Unfortunately, it does not appear that decision-makers can be relied upon to do so. The second defense of the potential Pareto criterion is that over time, policies will variously benefit and harm each group in society such that, on balance, the potential Pareto criterion will lead to actual Pareto improvements. The activists in the Occupy Wall Street movement would demur on
this point. Having rehearsed these obviously invalid arguments, Adler and Posner (2006), in their estimable book “New Foundations for Cost-Benefit Analysis,” conclude that the potential Pareto criterion “has zero moral relevance.”

The first of our criteria for evaluating policy analysis tools is that the information provided by policy analysis should be morally relevant. For this reason, our first recommendation is that welfare weights be applied to BCA. At the same time, we recommend that unweighted costs and benefit to groups at different income levels also be presented, as well as the estimated income level of each of the groups. There are three reasons for this. First, these numbers are necessary to compute welfare-weighted net benefit, and the computations should be made transparent. Second, some of the information contained in the unweighted benefits and costs will be relevant to decision makers for reasons other than welfare and equity, such as impact on government budgets or revenue impact on firms. Third, the unweighted net benefit of a policy is a necessary component of the five-step procedure we will recommend below.

We have noted that adjusting for diminishing marginal utility is a technical matter, and thus the decision of what weights to use should be made by analysts, based primarily on empirical estimates of the elasticity of marginal utility, also known as the Arrow-Pratt coefficient of relative risk aversion. We review the empirical literature below, along with a sample of the welfare weights implied by the various estimates. The bulk of these estimates converge on a relatively narrow range, allowing us to make a tentative recommendation about the weighting formula to be used. However, while the decision of what welfare weights to use clearly lies in the domain of analysts, the second of our criteria is that net-benefit estimates should be comparable across analyses. As such, we recommend that a determination of appropriate welfare weights for practical use be made by a centralized agency, such as the Office of Management and Budget in the U.S. or the Treasury in the U.K.
Finally, on a technical note, it is important to note than in order to apply welfare weights to the benefits and costs of groups with different levels of income, it is necessary to include financial transfers in the tally of costs and benefits to each group. For example, if a policy imposes a tax on a good that is largely consumed by the poor, only the deadweight loss from the tax is typically included in BCA. The tax payment by the poor is often ignored because it is exactly offset by the tax revenue received by the government. In order to apply welfare weights, both ends of the transfer must be accounted for, so that the full impact on the poor—deadweight loss plus tax payment—is tallied for the purpose of weighting. Likewise, for a policy that involves government use of labor, BCA would ordinarily tally only the opportunity cost of the labor (the area under the labor supply curve) but would ignore the wages transferred from the government to its employees. In order to apply welfare weights, both ends of the transfer must again be accounted for, so that the full impact on the employees—wages minus opportunity cost of labor—can be tallied for weighting purposes.

3.2 What to do about equity weights.

Before considering concrete recommendations about equity weights, it is important to understand the idea of plural values. Human beings place different kinds of value on different kinds of things, which is another way of saying that things matter for different reasons. In many cases, different kinds of value are considered incommensurable, a term whose definition has been the subject of much scholarship. Sunstein (1994) captures its meaning quite well: "Incommensurability occurs when the relevant goods cannot be aligned along a single metric without doing violence to our considered judgments about how these goods are best characterized." Roughly speaking, this essentially means that certain things cannot be measured on the same scale in a way that provides relevant or useful information about the ways in which those things matter.
Welfare and equity are incommensurable. They are different values that matter for different reasons. When anyone, anywhere, is made worse off, there is something bad about that, regardless of their income, and when anyone, anywhere is made better off, there is something good about that, because wellbeing and life satisfaction are of value, no matter who experiences them. However, when some people are very well off, while other people are much less well off, that is a bad thing, but for reasons that are very different. If the poor are made modestly better off, while the wealthy are made far better off, that is unambiguously good in terms of welfare, but it might be thought of as bad in terms of equity. The difference is even more stark if the gap between the welfare gain of the poor and the welfare gain of the wealthy is the result of some process that is deemed unfair. In contrast, if everyone was equally poor, it would be a bad thing in terms of welfare but might not concern us with respect to equity. We might even consider it to be a good thing with respect to equity, as many did under otherwise quite deleterious Stalinist regimes for much of the twentieth century.

Applying equity weights to BCA is an attempt to measure incommensurable values on a single scale. It is an attempt to say, "this state of affairs is better for society--taking all things into consideration--than that one because it scores higher on this cardinal scale." But doing so means the numbers no longer contain any information about either welfare or equity that is relevant to decision-makers. The fact that a certain policy is assigned a certain number by an equity-weighted BCA tells us nothing about how much welfare it generates and nothing about how it affects equity. The same number could be achieved by a policy that generates a great deal of welfare for relatively wealthy people while imposing a modest cost on the poor, or it could be achieved by a policy that closes the gap between the rich and poor while actually lowering

---

8 We are not suggesting that increasing the welfare of any given individual is a good thing overall, if, for example, that welfare is gained by unfair or nefarious means, or causes harm to someone else. We are simply saying that some positive value is generated, whether or not it is offset by a loss.
aggregate welfare. For example, the first policy might generate a welfare-weighted benefit to the wealthy of 100 and a cost to the poor of 25, a gain in aggregate welfare but a decrease in equity. The second policy might generate a welfare-weighted cost to the wealthy of 100 and a benefit to the poor of 75, a loss of aggregate welfare, but an increase in equity. With an equity weight of 2 on the welfare of the poor, both policies generate an equity-weighted net benefit of 50. A decision-maker learns literally nothing about the welfare and equity impacts of either policy from such equity-weighting, even if he or she understands that welfare and equity are plural and incommensurable values that matter to society for different reasons.

Adler (2008) makes the claim that "...distributively-weighted CBA provides guidance in balancing equity with overall welfare. The sum of weighted WTP/WTA amounts is meant to indicate whether, on balance, a policy should be pursued, given both distributive and aggregative considerations." In our view, exactly the reverse is true. Decision-makers are tasked with weighing the tradeoffs among disparate policy values, or what Bardach and Patashnik (2019) call "criteria," in order to make considered judgments about what is in the best interest of society “on balance.” Analysts are tasked with presenting information about those disparate criteria, that allows decision-makers to weigh those tradeoffs. Equity-weighted BCA literally prohibits decision-makers from weighing the tradeoff between welfare and equity. This brings us to our second criterion for evaluating policy analysis. Policy analysis should respect incommensurability and provide decision-makers with information that increases rather than decreases their ability to weigh tradeoffs among incommensurable values. Applying equity weights to BCA fails this test. For this reason, we recommend that equity weights not be used in BCA.

However, there is information contained in BCA that can help decision-makers weigh the tradeoff between welfare and equity, which is the breakeven weight that would change the sign of welfare-
weighted net benefits. If a policy fails welfare-weighted BCA because the welfare benefits to the poor are outweighed by the welfare costs to the wealthy, there will be some weight, greater than one, which, when multiplied by the welfare-weighted benefit to the poor, will cause the net benefit to society to be zero; above that number, the net benefit will become positive. (And vice versa when costs are to the poor, benefits to the wealthy, and the policy has positive welfare-weighted net benefits.) We call that number the “breakeven equity weight.” The breakeven equity weight is the ratio of the net impact (cost or benefit) on the wealthy, and net impact (benefit or cost) on the poor. For example, the first of the two policies would have a breakeven equity weight of $\frac{100}{25} = 4$, because the cost to the poor is one-quarter of the benefit to the wealthy, while the second policy would have a breakeven weight of $\frac{100}{75} = 1.33$, because the benefit to the poor is three-quarters of the cost to the wealthy.

The breakeven equity weight is a number that we believe to be intuitively comprehensible. Consider the first policy above, with a breakeven equity weight of 4. One interpretation of that number is that if we deem the welfare cost to the poor to "matter" at least 4 times more than the welfare gain to the wealthy, then the equity loss will outweigh the welfare gain. Meanwhile, in the case of the second policy, if we care about the welfare gain to the poor by at least 33% more than the cost to the wealthy, then the equity gain will outweigh the welfare loss. Our third criterion for evaluation policy analysis is that analysis should not involve moral judgments, but should empower decision makers to apply their own moral judgements. In the context of the two examples above, we believe that a thoughtful decision maker can consider where their own moral values fall with respect to the two breakeven equity weights, and that breakeven weights satisfy our third criterion.

This is not the same as measuring equity and welfare on the same scale. In a BCA, for example, welfare-weighted net benefit estimates may tell a decision-maker that a) there is net loss of welfare, and b) there
is a benefit to the poor and a cost to the wealthy. In other words, the BCA provides information about both welfare and equity. The breakeven weight allows the decision-maker to assess whether their own sense of the importance of equity is enough to outweigh the welfare impact. As such, it satisfies our incommensurability criterion. In effect, breakeven equity weights simply distill the information on distributional impacts inherently contained in a welfare-weighted BCA into a single number that is intuitively meaningful and useful.

The breakeven equity weight is only relevant when, after welfare weighting, there are either a) benefits to the poor and more than offsetting costs to the wealthy or b) costs to the poor and more than offsetting benefits to the wealthy. In these cases, there will be some number greater than one which, when multiplied by the welfare impact on the poor, will change the sign of welfare-weighted BCA. If there are benefits to the poor and less than offsetting costs to the wealthy, or costs to the poor and less than offsetting benefits to the wealthy, there is no weight on the welfare of the poor that will change the sign. Equity weighting would only make the case for the policy stronger or weaker as the case may be. Another way of putting it is that in either of these latter cases, welfare-weighted BCA alone provides adequate information about both welfare and equity for a decision-maker to see that the equity and welfare impacts of the policy move in the same direction.

It may be worth noting that if a policy generates benefits for both wealthy and poor, it could increase the gap between the wealthy and the poor. In other words, it could be better than the status quo in terms of welfare, but worse in terms of equity, even though it alleviates poverty. However, it is not possible to compute a breakeven weight, because inflating the benefit to the poor will only strengthen the case for the policy. If the benefit to the poor was modest and the benefit to the wealthy very great, one could imagine a decision-maker feeling the need to confront the tradeoff between the welfare and equity
consequences of the policy. Because there is no break-even equity weight in this case, a decision-maker would need to consult the benefits to the two groups in absolute terms to weigh this tradeoff.\footnote{If a policy generates costs to both poor and wealthy, it will reduce welfare, but could reduce the gap between wealthy and poor, if the cost to the wealthy is greater than the cost to the poor, and again, there is no way to compute a break-even equity weight. However, it is hard to imagine adopting a policy that hurts everyone in society strictly because it reduces the gap between wealthy and poor. Doing so would be like cutting off one’s nose to spite one’s face and would seem to require a rather high degree of schadenfreude towards the wealthy.}

3.3 What welfare weights to use?

As discussed in Section 2, determining welfare weights requires estimates of the marginal utility of income at different income levels, and in order to make this feasible, it is necessary to make an assumption about the functional form of utility. In the methodologies we rely on below, the standard assumption is that preferences can be represented by an isoelastic utility function:

\[ U(y) = \frac{y^{1-\epsilon}}{1-\epsilon} \quad \text{if} \quad \epsilon \neq 1 \quad \text{and} \quad U(y) = \ln(y) \quad \text{if} \quad \epsilon = 1 \] (1)

where \( y \) is income and \( \epsilon \) is the elasticity of marginal utility with respect to income.

Given this utility function, the marginal utility of income is:

\[ \frac{dU}{dy} = \frac{1}{y^\epsilon} \] (2)

Because \( \epsilon \) is the elasticity of marginal utility of income, with respect to income, it indicates the percentage decrease in the marginal utility of income caused by a 1-percent increase in income. The isoelastic function has the desirable property of constant elasticity, which allows for straightforward empirical estimation.

Given this utility function and an estimate of \( \epsilon \), welfare weights can be readily determined. For instance, setting the welfare weight for those at median income (the \( m^{th} \) group) equal to one, the welfare weight (WW) for a higher or lower income group (the \( i^{th} \) group) would be the ratio of the income for the \( m^{th} \) group to the income of the \( i^{th} \) group, taken to the \( \epsilon^{th} \) power:
The information necessary to compute \( \frac{y_m}{y_i} \) for U.S. households is readily obtainable from federal government statistics such as the annual U.S. Census Bureau’s Annual ASEC Survey, but computing welfare weights for various income groups also requires a value for \( \varepsilon \). A substantial number of estimates of \( \varepsilon \) from previous research can be found in Table 1. The table also shows the welfare weights implied by each estimate, which were obtained by using equation (3). The welfare weights pertain to the 25th, 50th, and 75th U.S. income percentiles in 2021. Because income is higher in larger households than in smaller households, and the average size of U.S. households is 2.5, the income amounts used to compute the welfare weights are a weighted average of the incomes of two-person and three-person households. As shown in Table 1, the weights must equal 1 at the 50th percentile because at that percentile \( y_m = y_i \) and hence \( \frac{y_m}{y_i} = 1 \). The table also briefly indicates the sort of data used to obtain each estimate.

Most of the estimates of \( \varepsilon \) in Table 1 are forced to be constant with respect to income. Thus, equation (3) can be used to readily compute welfare weights. If the elasticity varies with income, then the elasticity of the marginal utility of income must be determined at different income points along the utility curve. A few of the studies listed in the table test whether the null hypothesis of constant utility can be rejected, finding that it cannot be. The few estimates that allowed \( \varepsilon \) to vary with income imply that there is little change in \( \varepsilon \) as income changes. For these studies, we simply take the central tendency.

\[ w_i = \frac{dU_i}{dy_i} \frac{dU_m}{dy_m} = \left( \frac{y_m}{y_i} \right)^\varepsilon \] (3)

\(^{10}\) We recognize that determining the income distribution of those who experience either the costs or the benefits of the poor places an increased informational burden on the analyst. We address this briefly in the conclusion.
The researchers who estimate $\varepsilon$ also assume (often implicitly) that all persons have the same marginal utility of income function, so that interpersonal comparisons of utilities are possible. To the best of our knowledge, Demuynck (2018) provides the only test of the null hypothesis of homogeneous marginal utility of income, finding that it cannot be rejected.

The estimates of $\varepsilon$ in Table 1 are all for the U.S. and the U.K., two countries with similar cultures and economies, where one might anticipate that the elasticity values might be similar. Although estimates of $\varepsilon$ do exist for other countries, especially for those in Europe, the majority of work appears to be for the U.S. and U.K., especially for the latter. Much of the work in the U.K. seems to be motivated by the fact that government-funded project evaluation relies on a social discount rate, $d$, that is calculated by using the Ramsey equation (see H.M. Treasury 2018):

$$d = p + g\varepsilon$$

(4)

where $p$ is the rate of time preference and $g$ is the rate of growth of consumption.

Although the U.K. government has assumed that $\varepsilon$ equals one for purposes of determining $d$, as shown in Table 1, most estimates have found that it is greater than one, which has obvious implication for the discount rate and for welfare weights.\(^{11}\) In fact, in determining welfare weights, the U.K. Green Book (2018) uses a value of 1.3, which is taken from Layard, Mayraz, and Nickell. (2008)

\(^{11}\) When $\varepsilon$ is incorporated into the Ramsey equation and then used for discounting, this can be viewed as welfare weighting between the present and future. Declining discount rates, in contrast, can be viewed as equity weighting between the present generation and future generations. Thus, the former is based on empirical estimates, while the latter requires a value judgment.
As indicated in Table 1, a number of different approaches have been used in estimating \( \varepsilon \). Each of these approaches is briefly described in Appendix 1. Each approach has its own weaknesses and requires fairly strong assumptions. The best approach is not apparent to us. As discussed next, however, it is possible that the estimates in Table 1 tend to converge on what we might think of as the correct value.

The median value of \( \varepsilon \) in Table 1 is 1.48. Of the 21 values of \( \varepsilon \) listed in Table 1, 15 are between 1.2 and 1.7, a fairly narrow range. Three others are moderately larger at 2.0, 2.02, and 2.19. The three remaining values are outliers at 0.2, 0.3, and 3.57. The 1.2 value of \( \varepsilon \) implies that a typical household at the 25th percentile level, with about $43,664 in annual pre-tax income, would value an extra $100 by around twice as much as an average household at the median income level of about $77,267, while a household at the 75th percentile receiving about $123,000 in income would value the extra $100 by a little more than half as much. The 1.7 value for \( \varepsilon \) implies that a typical household at the 25th percentile income level would value the $100 by more than two-and-a-half times as much as a household at the median, while an average household at the 75th percentile would value the $100 by around 40 percent as much as a household at the median income level.

Using a few meta-analytic tools (see Shadish and Haddock, 1994 and Hedges, 1994), we pooled the estimates of \( \varepsilon \) in Table 1 and estimated their mean value. In calculating a mean value, meta-analysis takes account of the fact that some estimates are more statistically precise than others, as implied by their smaller variances, and calculates a weighted mean, the weight being the inverse of the variance of each of the individual estimate of \( \varepsilon \). By using this weighting scheme, estimates of \( \varepsilon \) that are of greater statistical significance contribute more to the pooled mean than estimates that are less so.
The weighted mean of $\varepsilon$ is 1.51, with a small standard error, 0.032, and a tight confidence interval of 1.45-1.58. For comparison, the unweighted mean is 1.58. These means are based on 14 of the 21 estimates in Table 1. If all 21 estimates are included, the unweighted mean drops slightly to 1.50. The variance was not available for four of the six estimates that rely on the progressivity of tax rates and for Demuynck (2018). We excluded the two outlier estimates of 0.2 and 3.57, as the first of these estimates had an extremely small variance and thus a very large weight in the analysis and the second had an exceptionally large variance and thus a very small weight. In fact, the weight of the former is almost 15,000 times larger than the weight of the latter. (The median weight is 37.5, with most of the remaining weights between 2 and 100.) With the two outliers, the correlation between $\varepsilon$ and the weight is -.59; but without them, it is only -.14. Inclusion of the two outlier estimates would have caused the weighted mean to fall well below one.

We had particular concerns about the two estimates of $\varepsilon$ that are based on the progressivity of tax rates approach because we believe that that the determination of tax rates, which is done by politicians, depends on many influences other than the utility that different income groups receive from additional dollars—for example, equity concerns, concerns over how tax rates affect work and investment incentives, and the relative strengths of different interest groups. However, if we exclude these two estimates of $\varepsilon$, the weighted mean changes only very slightly from 1.51 to 1.57, with a confidence interval of 1.46 to 1.69.

It is important to stress that our meta-analysis is preliminary. We strongly recommend that a more thorough meta-analysis be conducted. Although we believe that we have a good sampling of estimates of $\varepsilon$ appearing in published articles, such a study would likely find additional values for inclusion, especially those that have not been published. In addition, the various methodologies used to derive estimates need

12 The variance is not available for the other outlier, Demuynck (2018).
a careful review. Weighting each estimate for its quality, as well as for its variance, might then be possible (see Shadish and Haddock, 1994). Considering these factors, we very tentatively conclude that it is reasonable to base welfare weights on an elasticity of 1.5, with sensitivity testing at 1.3 and 1.7.

3.4 Summary of recommendations.

In summary, our recommendations to analysts conducting BCA are these:

1. Compute and present both unweighted and welfare-weighted net benefits, in dollars, being careful to include financial transfers in the costs and benefit to different groups. Ideally, the welfare weights that are used should be those provided by a central agency. If such values are not available, they can be based on estimates of \( \varepsilon \) selected from the central range of estimates found in the literature, such as those appearing in Table 1. The mean value for \( \varepsilon \) and values for sensitivity testing should rely on a meta-analysis, such as the one we have conducted.

2. Explain the concept of welfare-weighted net benefits. Explain the difference between potential Pareto efficiency and welfare efficiency, and comment on the moral relevance of the two measures.

3. Present welfare-weighted costs and benefits by group and point out the equity implications of the distribution of costs and benefits.

4. If, after welfare-weighting, there are benefits to the poor and more than offsetting costs to the wealthy, or costs to the poor and more than offsetting benefits to the wealthy, compute the breakeven equity weight. The BCA report should explain how these breakeven weights should be used. Note that this step can be applied to dimensions of distribution other than income, such as minority/non-minority or male/female.
Because both welfare weighting and equity weighting are controversial and may be difficult for decision-makers who are unfamiliar with the concepts to fully comprehend, we suggest a stepwise method for making use of these tools that minimizes the extent to which challenging and controversial concepts need to be introduced and explained.

First, compute unweighted net benefits.

Second, if there are costs to the poor and the total net benefit is negative, or there are benefits to the poor and the total net benefit is positive, stop. In the first instance the status quo dominates the policy and in the second instance the policy dominates the status quo, with respect to both equity and welfare. Welfare weighting is unnecessary for determining whether or not to adopt the policy. A possible exception occurs if there are modest benefits to the poor and much larger benefits to those better off. In that case, the policy would cause the gap between the two groups to widen, which would be a decrease in equity, that could potentially be seen as outweighing the welfare benefits of the policy. In this case, the determination of whether, on balance, the equity impact outweighs the welfare impact would need to be made by direct inspection of the benefits to each group.

Third, if the second step does not hold, apply welfare weights, being careful to account for financial transfers.

Fourth, if the sign of net benefits changes, stop. The policy dominates the status quo with respect both to equity and welfare.

Fifth, if the fourth step does not hold, compute the breakeven equity weight.
In addition to these steps, we feel that it is valuable, in all cases, to present welfare-weighted costs and benefits by group in a table of impacts. This allows decision makers to directly inspect the welfare and distributional implications of the policy. Such a table would look as follows:

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Wealthy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits</td>
<td>$B_p$</td>
<td>$B_w$</td>
<td>$B_p + B_w$</td>
</tr>
<tr>
<td>Costs</td>
<td>$C_p$</td>
<td>$C_w$</td>
<td>$C_p + C_w$</td>
</tr>
<tr>
<td>Net</td>
<td>$B_p - C_p$</td>
<td>$B_w - C_w$</td>
<td>$(B_p + B_w) - (C_p + C_w)$</td>
</tr>
</tbody>
</table>

For a policy that generates costs to the poor and benefits to the wealthy, we would have $B_p = C_w = 0$. The analyst would then report the welfare impact, $B_w - C_p$, the equity impact, $B_w + C_p$ (the amount by which the policy increases the welfare gap between wealthy and poor), and the breakeven weight, $\frac{B_w}{C_p}$. This presentation provides all of the information available in BCA about the tradeoff between welfare and equity.

3.5 Illustrating the steps.

In this section, we illustrate how the steps can be used in practice. Our three examples are drawn from a review of 26 BCAs of welfare-to-work programs (Boardman et al., 2018), which were evaluated by random assignment by MDRC, a New York-based evaluation firm. The programs provided information that was helpful for deciding whether the programs should continue to operate. Because these programs were all targeted at single-parent welfare recipients, distributional impacts are of obviously especial importance. The BCAs of these programs divide those affected into two groups: participants who received services
under the programs, and non-participants, a category that includes everyone else in society. The key BCA components for participants were earnings, which were usually positive, and welfare receipts, which are typically negative; the major components for non-participants were the taxes required to operate the program and reductions in the taxes needed to support the welfare system. Benefits and costs that were estimated for both participants and non-participants were divided by the number of participants and are thus comparable.

Because the programs are small relative to the population, the incomes of non-participants should be close to the national median. For purposes of the illustration, the median incomes of participants are assumed to be at the 25th percentile, which is probably a reasonable approximation, given that participants entered the program when they were on welfare. The findings from the meta-analysis describe earlier suggested setting the elasticity of the marginal utility of income, $\varepsilon$, at 1.5 in conducting welfare weighting, which implies a welfare weight of 2.35 at the 25th percentile level and a weight of 1 at the 50th percentile level.

Our first example pertains to the GAIN program, which, beginning in 1987, operated in Riverside, California. The BCA of this program estimated that the unweighted net benefits of participants and non-participants in 2021 dollars were $2,951 and $5,718, respectively, per participant. Total unweighted net benefits were obviously positive at $8,669 per participant. As indicated by the second step, welfare weighting is unnecessary because both participants and society as a whole received positive net benefits, implying that the policy passes the BCA test on the basis of potential Pareto efficiency, so that, for the purpose of deciding whether to adopt the program, welfare weighting is not necessary. The unweighted net benefit number indicates that the program dominates the status quo with respect to both efficiency (both potential Pareto efficiency and welfare efficiency), and equity. Indeed, welfare weighting is often
not required in practice. For example, in 16 of the reviewed 26 BCAs, including Riverside GAIN, there were either costs to the participants and the total unweighted net benefit was negative or there were benefits to the participants and the total unweighted net benefit was positive.

The second example is for a so-called jobs first program, which began in 1991, and also operated in Riverside, but less successfully than Riverside GAIN. Participants had estimated negative unweighted net costs of $2,238 in 2021 dollars, although because of reductions in welfare payments, non-participants had estimated unweighted net gains of $2,605 per participant. Consequently, there were small total net benefits of $368. As indicated by the third step, welfare weighting is necessary for assessing the program because there are costs to participants, yet total net benefits are positive. Not surprisingly, given the small size of the estimate of total net benefits, they become negative—at $2,653—once they are welfare weighted. Given the sign flip, the fourth step implies that the fifth step, explicit consideration of equity, is unnecessary. With welfare weighting, the status quo is superior with respect to both welfare and equity.

Our final example is for a welfare-to-work program in which equity considerations are quite relevant: a GAIN program that beginning in 1988 operated in Alameda, California. Like Riverside GAIN, this program was one of the GAIN programs that operated in each California county towards the end of the last century, but it was administrated under a different philosophy than Riverside GAIN. Like Riverside GAIN, participants were found to be better off by the BCA, with unweighted net benefits estimated at $1,698 in 2021 dollars, but unlike Riverside GAIN, the BCA also found that there were unweighted net costs to non-participants of $5,788 per participant. Consequently, the unweighted loss to society was estimated to be $4,090, a reduction in potential Pareto efficiency. In this case, as per step 3, welfare weighting is necessary, and causes net benefits for participants to increase to $3,990, net costs for non-participants remain at $5,788 per participant, and total net costs to shrink to $1,798, a reduction in welfare efficiency.
Because there are no sign changes, it is necessary to move on to the fifth step and compute the breakeven equity weight. Based on the weighted estimates, the breakeven equity weight is 1.45, suggesting that to approve the program on BCA grounds, the decision-maker would need to value any given amount of utility received by a person at the 25th percentile by more than 45 percent of the same amount of utility received by someone at the national median *because* it improves equity.

We feel that calculating a breakeven equity weight after applying welfare weights is intuitively compelling. We believe decision-makers will be able to comprehend the distinction between bias in WTP causes by diminishing marginal utility of income and the possibility that society might simply care more about the welfare of the poor than the wealthy. However, we acknowledge that there is another approach to thinking about redistributive policies, which is to ask, “How much more valuable is it, to society as a whole, all things considered, to give a dollar to a poor person than to give the same dollar to a wealthy person?” The question rolls together the idea that money matters more to the poor, and the idea that the poor matter more to society. For a decision-maker who thinks in this way, it might be more helpful to skip welfare weighting and simply compute the breakeven weight on the unweighted costs and benefits of the policy. For example, in the case of the Alameda GAIN program, the breakeven weight on the unweighted costs to participants is 3.41, implying that after taking both welfare and equity into account, the decision-maker would have to value each dollar received by participants at over three-times a dollar received by non-participants.

As part of a multigoal analysis, the decision-maker would have other concerns besides the distribution of costs and benefits among income groups. For example, in the Alameda GAIN program, operating costs were estimated to be $10,893 per participant, in 2021 dollars, which is large for welfare-to-work programs. The decision-maker would presumably multiply that figure by the expected number of future
participants to determine if there is sufficient space in the government’s budget. As this discussion suggests, while welfare weighting is helpful, unweighted estimates of benefit and costs are also needed for decision making.

**Section 4: Conclusion**

In brief, our thesis is two-fold: First, using welfare weights to correct for diminishing marginal utility of income is a technical solution to what is essentially a methodological problem: WTP is the only viable metric for BCA, but it is a biased measure of welfare. Without welfare weights, BCA is an implementation of the potential Pareto criterion, and measures what we call “potential Pareto efficiency.” However, the potential Pareto criterion has been shown to have no moral relevance. Our first criterion for developing our recommendations is, *the information provided by policy analysis should be morally relevant*. To make BCA morally relevant, we favor correcting for bias using welfare weights, and viewing BCA as a measure of aggregate welfare, or what we call “welfare efficiency.”

The decision of what welfare weights to use is a technical decision, rather than a moral judgment, and consequently belongs under the purview of the analyst. However, in order to maintain comparability across analysts and analyses, some central body should determine the correct weights, on the basis of empirical evidence. We have used meta-analysis to generate specific recommendations regarding correct weights, which we believe could be used with some confidence by such a central body.

There may be those who feel that welfare-weighted BCA is not actually BCA anymore. We believe this position is rooted in a misunderstood conflict between the potential Pareto definition of BCA and the aggregate welfare definition of BCA. As we have demonstrated, BCA cannot be both an implementation of the potential Pareto criterion and a measure of aggregate welfare. We must choose between the two.
Welfare-weighted BCA is not a measure of potential Pareto efficiency. Consequently, for those who cleave to the potential Pareto definition of BCA, welfare-weighted BCA is not actually BCA anymore. This is correct if BCA is defined in terms of the potential Pareto criterion. However, a policy that satisfies the potential Pareto criterion may reduce aggregate welfare, and in particular, may lower the welfare of the poor more than it increases the welfare of the wealthy. This is simply an unacceptable decision-making criterion. The idea that we should adopt a policy simply because the wealthy could, if they chose, or were compelled to, pay off the poor, serves, among other things, as a justification for making the wealthy better off at the expense of the poor.

This leaves us with the definition of BCA as a measure of aggregate welfare, a decision-making criterion that is entirely valid in the context of a multi-goal analysis in which other considerations are assessed, including distributional impacts. Under this definition of BCA, it is the unweighted net benefit that is not actual BCA, because it does not measure welfare. Under the aggregate welfare definition of BCA, WTP must be weighted by the relative marginal utility of income. To those who consider this to be an unacceptable deviation from what BCA is supposed to be, we simply emphasize that support of the unweighted net benefit demands a definition of BCA as something that is morally invalid. If the argument is then made that unweighted BCA, though biased against the poor, is an adequate approximation of aggregate welfare, our response is, why not eliminate the bias and make the measure correct rather than approximate, as there is a straightforward, comprehensible, and empirically justifiable methodology for doing so?

At first glance, it might appear that welfare-weighted net benefit no longer has any cardinally meaningful units. It is ostensibly denominated in dollars, but because WTP is weighted by a number based on the marginal utility of income, and because, for any given set of preferences, there is an infinite number of
marginal utility numbers, these so-called dollars are no longer meaningful as a cardinal scale. This view is based on a misunderstanding of the way that welfare weights are calculated. They are the ratio of the marginal utility of income at median income and the marginal utility at any other level of income, and since utility functions are unique up to an affine transformation, using this ratio means that there is only one set of welfare weights for any given set of preferences. Welfare weights convert the dollar value of a given quantum of welfare to a person of any given income level to the dollar value of the same quantum to a person with median income. We believe welfare-weighted net benefit is an intuitively comprehensible number: it is the net benefit to society, across income levels, expressed in terms of the monetary value of utility to a median-income household. In other words, it is the net benefit we would measure if all households had incomes at the median.

Our second thesis is that the use of equity weights in BCA, in an attempt to account for what has been referred to as “diminishing moral value of welfare,” (Adler 2008) should be completely avoided. Equity weights render BCA fundamentally uninformative, because equity weighted net benefit contains no information about either welfare or equity as distinct, incommensurable values. The same equity-weighted net benefit can be generated by a policy that increases aggregate welfare while decreasing equity, or by a policy that decreases aggregate welfare while increasing equity, or by any combination in between. Confronted with equity-weighted net benefit, decision-makers can draw no conclusions whatsoever about either the welfare impacts or the equity impacts of a policy. Our second criterion for developing our recommendations is, policy analysis should respect incommensurability and provide decision-makers with information that increases rather than decreases their ability to weigh tradeoffs among incommensurable values. To satisfy this criterion, equity weights must not be used. Nonetheless, intuitively useful information about equity can be extracted from a welfare-weighted BCA by calculating the breakeven weight on the poor that changes the sign of net benefits from positive to negative, or vice
versa. This number can be thought of as an indicator of how much more a decision-maker would need to value the welfare of the poor than the welfare of the wealthy, in percentage terms, in order to conclude that either a) a policy that benefits the poor but reduces aggregate welfare should nonetheless be implemented on the basis of equity, or b) a policy that harms the poor but increases aggregate welfare should nonetheless not be implemented on the basis of equity. We believe that this is an intuitively comprehensible number, and satisfies our third criterion for developing our recommendations, which is, 

*analysis should not involve moral judgments, but should empower decision makers to apply their own moral judgements*

We have provided concrete, stepwise recommendations for how welfare weights should be applied, how breakeven equity-weights should be used, and how the results should be communicated to and used by decision-makers. Our hope is that these recommendations will be adopted by central agencies such as the Office of Management and Budget in the U.S., the Treasury in the U.K., or the European Commission in the E.U.

Some defense may be called for of the position that policy analysis should provide information about incommensurable values separately, so that decision-makers can weigh tradeoffs among incommensurables according to their own values. An alternative position, expressed by Adler (2008), is that policy analysis should “provide guidance” on what decisions to make. The claim is made that by generating an ostensibly cardinal measure of what is sometimes referred to as “social welfare,” the application of equity weights tells decision-makers what policies to adopt, once welfare and equity have been (quite literally) weighed against one another. In our view, this is not the role of the analyst. The

13 The claim could be made that equity-weighted BCA is not actually Social Welfare Measurement, but Adler (2013) himself has stated that equity-weighted BCA “mimics” a particular social welfare function. This is close enough to defining equity-weighted BCA as a form of social welfare measurement for our purposes. Certainly equity-
application of moral valuation to public decision making is the role of democratically accountable decision-makers, not analysts. Rather, the role of the analyst is to inform such decision-makers so that they can apply their own moral valuation when weighing the different incommensurable impacts of a policy. The claim could be made that if a decision-maker agrees with the equity weights that are used in any given analysis, they can rely upon equity-weighted BCA to capture their own moral valuation. To see why this is not valid, consider that decision-makers must also weigh welfare and equity impacts against other considerations, such as rights, fairness, political feasibility, and budgetary impact. Equity-weighted BCA is useless for weighing those tradeoffs because, as we have demonstrated, it contains no information about either welfare or equity separately.

The idea that policy analysis should provide information rather than guidance is related to a concern about the recent Biden memorandum. The memorandum calls for “recommendations [that] provide concrete suggestions on how the regulatory review process can promote public health and safety, economic growth, social welfare, racial justice, environmental stewardship, human dignity, equity, and the interests of future generations.” (Emphasis added.) It is not entirely clear what the authors of the memorandum have in mind, but we are concerned by the idea that policy analysis, by itself, should mechanically promote any particular kind of policy. If “promoting” equity simply means calling attention to the equity implications of public decisions, that is appropriate in our view. Perhaps it is felt that if attention is not drawn to equity impacts, then welfare impacts will inevitably dominate decision making. If that is the concern, breakeven weights solve the problem—by drawing attention to equity impacts—while equity-weighted BCA does not—because it actually provides no information about equity. Rather, we view equity-weighted net benefit can no longer be defensibly thought of as being denominated in dollars in a meaningful way. The units of social welfare do not represent the value of policy impacts to individuals of any income level. They represent the moral value of those impacts to a decision-maker who happens to hold moral values consistent with the particular social welfare function being used.
weighted BCA as a way of mechanically promoting equity. If decision-makers treat equity-weighted BCA as a measure of some meaningful, single, or “monistic” moral value (which, as a cardinal measure of incommensurable values, it is not), they may be tempted to treat a positive equity-weighted “net benefit” as a strike in favor of a policy, without any understanding of what that strike actually represents, i.e., whether it means the policy has positive welfare impacts, positive equity impacts, or some combination of the two. We strongly recommend that the White House clarify what is meant by “promoting equity,” and that it adopts our interpretation of providing information about equity rather than mechanically tipping the balance in favor of equity.

We acknowledge that our approach to distributional weight has some weaknesses. Perhaps foremost is that in some cases, it will impose a high burden upon the analyst. In particular, the distribution of impacts across income groups must be known. In some cases, such as the welfare-to-work programs we discuss in section 3, the necessary information will have been collected in the course of conducting a program evaluation. In some cases, it may be relatively straightforward to approximate the income distribution of impacted groups. For example, consider a policy of taxing cigarettes to fund early-childhood education. The income distribution of smokers could be computed using the National Survey on Drug Use and Health, and the income distribution of families receiving early-childhood education support could be approximated by looking at the income distribution of household with children of pre-school age at or below the income threshold for the program. Some cases may present greater, but not unsurpassable challenges. Consider a workplace-safety regulation that benefits workers in a particular occupation at the cost of corporate profits. The income distribution of workers in the affected occupation could be estimated using data from the Occupational Employment and Wage Statistics program of the Bureau of Labor Statistics. Estimating the income distribution of corporate shareholders might be more challenging, but some kind of approximation might be possible, using assumptions no stronger than those often used
in other, more conventional aspects of BCA. Additionally, there may be cases in which a cogent argument can be made that there are no significant distributional impacts to consider. Constructing a bridge that connects two metropolitan areas, for example, is likely to generate benefits and costs that are relatively broadly distributed across the income distribution of the region. Finally, there are impacts that are common across a broad range of policies for which the income distribution could be estimated by scholars or some central analysis agency. For example, the income distribution of payers of different types of taxes in different jurisdictions could be estimated in much the same way as the Marginal Excess Burden of Taxation, which has been estimated by scholars, whose estimates are widely used by analyst. Finally, if there is no basis for approximating distributional impacts, analysts can address concerns non-quantitatively in much the same way that is recommended for non-quantifiable or non-monetizable impacts conventionally.

A related weakness is that our approach to welfare weights is based on empirical estimates of the marginal utility of income, but these estimates could conceivably be non-trivially off the mark. Perhaps even more importantly, what really matters is not the income level of any given individual, but their actual consumption, which may include intangible “goods” such as health and fatality risk. Arguably, an additional dollar has greater value, ceteris paribus, to an individual in poor health than to a healthy individual, and the same may be true with respect to urban amenities, pollution, or even a felicitous work environment. The argument would be that as utility from total consumption goes up, the value of the additional utility that can be attained by spending a dollar on additional consumption goes down. If that is correct, we would need a measure of the marginal utility of consumption, rather than income, and we would need to measure all dimensions of consumption in order to determine where in the distribution of

---

14 Though this determination would depend upon consideration of the most common users of the bridge. The Golden Gate Bridge, in the San Francisco Bay Area, for example, connects two quite affluent communities. Here again, however, some approximating assumptions about the income distribution of bridge users might not be out of reach.
consumption any given individual falls. This is obviously not practicable. Perhaps we can take refuge in the idea that non-purchased dimensions of consumption are likely to be positively correlated with income, and that, because welfare-weights are relative to median income, they are likely to be roughly correct with respect to total consumption. This is probably not empirically verifiable, but we feel that the income-based approach to welfare weights is an important step in the right direction.

In our recommendations, we have not addressed the issue of how to rank multiple policies with respect to equity. In Appendix 2, we consider the case of two policies compared to the status quo. Each policy can be compared to the status quo using our five-step process. If both policies are deemed superior to the status quo, comparing the two to one another requires a modified approach, as it is not possible, in general, to compute breakeven equity weights between the two policies. In this case, the information contained in tables of impacts can be used to weigh the tradeoffs between efficiency, equity, and, in the two policy case, poverty alleviation.

Likewise, our recommendations do not address the issue of comparing policies that have impacts on more than two groups. We have worked out a straightforward way to compare the equity impacts of policies with impacts on three groups. Consider a policy that has different impacts on the poor, the wealthy, and those of median income. For any given equity weight on the poor, a breakeven equity weight can be computed for the median, and vice versa. This makes it possible to compute what we refer to as a “locus” of breakeven equity weights, which can be presented graphically. The locus allows decision-makers to simultaneously weigh the impacts of the poor and the median against the wealthy. For example, a decision-maker can first consider their own threshold equity weight on the poor and then use the breakeven locus to identify the corresponding breakeven equity weight on the median. If that weight is acceptable to the decision-maker, then he or she can safely conclude that the equity impacts of the policy
overall are acceptable to them. The breakeven locus for more than three groups has three or more
dimensions and is thus probably not presentable in a comprehensible way. Additionally, there are
situations in which the breakeven locus would involve higher equity weights on the median than on the
poor, in which case breakeven weights would be meaningless. We provide worked examples in Appendix
3.

Breakeven equity weights can be computed for dimensions of distribution in addition to income, such as
race or gender, though here the information burden is even greater because distribution by income must
first be determined for the computation of welfare weights, and then welfare-weighted costs and benefits
must be identified by race or gender. Here too, a locus of breakeven weights could be used across as many
as three groups, but not more.

Throughout this paper, and throughout the literature on distributional weighting, the term “equity” has
been used in a very restricted sense, as a synonym for “distribution.” A policy is either equitable or
inequitable to the extent that it either decreases or increases the gap between the welfare of one group
and the welfare of another. Perhaps a better term for this would be “equality of welfare outcomes.” The
word equity is also used to capture a range of other policy concerns: equal opportunity, fairness, justice,
and perhaps even dignity. We have no recommendations for how to account for these additional
dimensions of equity, though understanding distributional impacts is likely to be a necessary ingredient
in any such analysis.

To conclude, we see our paper as a direct response to the clarion call of the Biden White House for
concrete recommendations on ways to modify “regulatory review” to account for equity. We have delved
into philosophical and theoretical considerations, in order to defend our position on welfare and equity
weights. But this is not a philosophical or theoretical paper. It is a paper about how to conduct BCA so that it is morally relevant, and so that it provides intuitively comprehensible information about the incommensurable values of welfare and equity. Our approach is intended to prevent BCA from being seen as irrelevant. On the one hand, the Biden memorandum makes it clear that if BCA does not begin to confront its inherent bias against the poor, it will become irrelevant in the eyes of those who take welfare seriously. At the same time, if equity weights are applied, and the information contained in BCA about incommensurable moral values is lost, there is a risk that BCA will be seen as irrelevant by those who take seriously the provision or comprehensible and useful information to decision makers. It is our hope that our paper will receive attention from decision-makers within the U.S. federal government and beyond.
REFERENCES


Johansson-Stenman et al. (2002)


U.S. Census Bureau’s Annual ASEC Survey in September 2021


Competing interests: The authors declare none.
Appendix 1: Methods for estimating the elasticity of the marginal utility of income.

**Convex Time Budget Experiments.** This approach, which is relatively new and so-far not widely used, involves a laboratory experiment in which subjects allocate a pot of money between two time periods, with an interest rate applied to any money allocated to the later time period. Regressing the allocation amounts on the interval between the earlier and later periods and the interest rate, allows the discount rate and the curvature of the utility function to be simultaneously determined, under the assumption that utility can be modeled as a power function. The curvature of the utility function, in turn, provides an estimate of ε. The methodology presents a trade-off for researchers. In order to achieve incentive compatibility, researchers must use real stakes, which requires that stakes be relatively low (for budget reasons) and that time horizons be short (for practical reasons). In order to elicit preferences using larger stakes and longer time horizons, which would be more ideal, it is necessary to use hypothetical stakes, which are considered less reliable.

**Indirect behavioral evidence: Consumer Demand Analysis.** This approach focuses on how utility is affected by the consumption of a particular good or broad category of goods (e.g., all foods). Essentially, the estimate of ε is the ratio of the good’s income elasticity to its own price elasticity, with an adjustment for the share of the total budget accounted for by the good. Findings may, of course, be sensitive to the specific specification of the regression model used to estimate ε. Note that of all the estimates of ε in Table 1, Demuynck (2018) is the only one that relies on a nonparametric conditional mean estimator instead of assuming isoelastic utility. The approach depends strongly on the assumption that the particular good selected is additively separable (i.e., additional utility from consuming the good does not depend on the quantity of any other good that is consumed). Thus, the good selected is often “all foods.” Still, there is no guarantee that the assumption is met, although some researchers have tested it and found it valid.

**Indirect behavioral evidence: Lifetime consumption models.** Many attempts have been made to estimate ε by using micro or macro time series data. In doing this, researchers often use a log-linearized consumption Euler equation, regressing consumption growth (g_t) on the real rate of return on assets (d_t): \( \ln g_t = c + bd_t + e_t \), where c is the constant, coefficient b is the inverse of ε, and \( e_t \) is the error term. As can be seen, the regression that is used bears a close kinship with equation (4), the Ramsey equation, once equation (4) is divided by ε. The estimation requires a number of assumptions, arguably the important of which are that capital markets are perfect and that the utility function is additively separable. The estimates may be biased if periods of market turbulence or changes in financial regulations occur. They may also be sensitive to which goods are defined as consumption goods.

**Indirect behavioral evidence: Relative risk aversion models:** Researchers have attempted to use insurance data to estimate the so-called coefficient of risk aversion, a measure of willingness to pay to avoid risk to incomes. The coefficient of risk aversion is assumed to correspond to the elasticity of the marginal utility of income. For purposes of the model, it is assumed that the amount of insurance increases with wealth but declines with insurance loading—that is, to the extent insurance total premiums exceed the total value of claims. Insurance loading become less important, the more important risk aversion is. In the estimation model, risk aversion is assumed to be constant, which is equivalent to assuming that ε is constant. The model presumes a rational insurance market in which consumers have knowledge of the probability that they will make a claim.

**Relating subjective wellbeing to income.** A number of happiness surveys have been conducted in recent years in which respondents rate their life satisfaction or “happiness” on a scale of 1-5 or 1-10. To use these findings to estimate ε, analysts must first make an assumption about the functional relationship between
the reported life satisfaction of individuals and their utility. That is, happiness is linearly linked to utility through a transformation: \( h_t = u_t + v_t \), where \( h_t \) is the measure of happiness for individual \( I \) at time \( t \) and \( v_t \) is an error term that is independent of the circumstances affecting utility (see Layard, Mayraz, and Nickell (2008). Given this transformation, and hence a measure of the utility of the surveyed respondents, \( \varepsilon \) can be estimation through a modification of equation (1). This approach appears to have been used only once (Layard, Mayraz, and Nickell, 2008), but in a frequently cited paper.

Direct elicitation. One way to determine the value of \( \varepsilon \) is to survey persons who are familiar with the concept, an approach that to the best of our knowledge has been used only once (Drupp, 2018). It is not clear how even experts could knowledgeably guess at the value unless they were aware of at least a few of the studies appearing in Table 1 or had values in mind for \( d, p, \) and \( g \) and used equation (4) to obtain a value for \( \varepsilon \).

Progressivity of income taxes. An estimate of \( \varepsilon \) can be obtained by using progressivity in income tax schedules under an assumption of equal sacrifice. In other words, that society’s aversion to inequality is such that income tax rates are set so that the loss in welfare is equal among taxpayers at different income levels. Operationally, this method, which has been frequently used, involves estimating the following equation: \( \varepsilon = \ln(1 - MTR)/\ln(1 - ATR) \), where MTR is the marginal tax rate and ATR is the average tax rate (Evans, 2004). This approach abstracts from the possibility that in addition to equal sacrifice, tax rates are influenced by concerns about inequalities in the income distribution and potential disincentive effects, as well as the relative strengths of different interest groups.

Appendix 2: Comparing two policies plus the status quo on the basis of equity.

Our recommendations in the main text involve the comparison of one policy to the status quo. In this appendix, we consider the comparison of two policies (A and B) that are alternatives to the status quo. First, apply the five-step procedure outlined in the body of the paper to compare each policy to the status quo. There are three possibilities: the status quo is preferred to both policies, the status quo is preferred to one policy but not the other, both policies are preferred to the status quo. In the first two cases, no comparison between A and B is necessary. In the third case, with the status quo ruled out, the comparison of A and B becomes relevant.

Unlike the case of comparing one policy to the status quo, in general it is not possible to compute breakeven weights between two policies. Depending on the valence of impacts on the poor and the wealthy (i.e., who gets the costs and who gets the benefits), there are cases in which a breakeven weight would have to be less than one, or even less than zero, which of course, is meaningless.

However, we can still extract information about welfare and equity from a table of welfare-weighted impacts, by group. Compute the welfare-weighted total net benefit of each policy, and the increase or decrease in the gap between wealthy and poor for each policy. For example, if \( C^A_p, C^B_p, B^A_w, \) and \( B^B_w \) are the costs to the poor and benefits to the wealthy of the two policies--after welfare weighting--total net benefits are \( B^A_w - C^A_p \) and \( B^B_w - C^B_p \), respectively, while the amount by which the two policies increase the welfare gap between wealthy and poor is \( B^A_w + C^A_p \) and \( B^B_w + C^B_p \), respectively. (For a policy that generates benefits to the poor and costs to the wealthy, the sum of benefit and cost for each policy is the amount by which the two policies decrease the welfare gap.) Unlike the case of comparing one policy to the status quo, there is an additional criterion to weigh against welfare and equity, which is the absolute impact on the poor, which can be thought of as the “poverty effect.” It is possible for A to increase (decrease) the
gap between wealthy and poor by less(more) than B while simultaneously generating a greater(smaller) absolute cost(benefit) for the poor. For example, suppose A has cost to the poor of 5 and benefit to the wealthy of 30, while B has cost to the poor of 10 and benefit to the wealthy of 20. A is superior with respect to welfare, but inferior with respect to equity (the amount by which it increases the welfare gap between wealthy and poor), but it imposes a smaller absolute cost on the poor. Thus, there are three criteria for a decision maker to consider.

One possibility is that one of the policies will dominate the other with respect to all three criteria. If not, the tradeoffs among the three criteria must be considered, in precisely the way that incommensurable criteria must be considered relative to one another in any multi-goal analysis. One policy may be superior to the other on any two of the three criteria. For example, policy A could generate greater welfare than B, while increasing the equity gap by less than B, and thus dominate with respect to welfare and equity, while generating greater absolute cost to the poor.

As with the case of one policy compared to the status quo, in the body of the paper, we recommend that a table of welfare-weighted costs and benefits be presented for each policy, and the welfare, equity, and poverty impacts of each policy presented in a third table. For the example above, the presentation would be as follows:

Table of impacts for policy A.

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Wealthy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits</td>
<td></td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Costs</td>
<td>−5</td>
<td></td>
<td>−5</td>
</tr>
<tr>
<td>Net</td>
<td>−5</td>
<td>30</td>
<td>25</td>
</tr>
</tbody>
</table>

Table of impacts for policy B.

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Wealthy</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits</td>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Costs</td>
<td>−10</td>
<td></td>
<td>−10</td>
</tr>
<tr>
<td>Net</td>
<td>−10</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Welfare, equity and poverty comparison.

<table>
<thead>
<tr>
<th></th>
<th>Policy A</th>
<th>Policy B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare (net benefit)</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Equity (increase in gap)</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Poverty (cost to poor)</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

It is worth noting that this approach is relevant when the two policies generate benefits to both groups. As with the case of one bearing the cost and the other group receiving the benefits, the tables above can be presented. Any combination of welfare, equity and poverty comparisons is possible, and the tradeoffs among the three can be considered.
This approach can be applied to any number of policies compared to the status quo, though the demand upon the cognitive capacity of the decision maker increases with the number of policies under consideration.

Appendix 3: analysis of policies with more than two groups.

Here we develop an approach to breakeven equity weights that can be applied to policies that have differential impacts on more than two groups. Conceptually, this can be done for any number of groups, but it is only practicable for three groups. We will consider policies that have impacts on the poor, the wealthy, and those of median income. For any given policy, the impacts on each group are tallied, and welfare weights applied. Then, what we call a “locus” of breakeven weights is computed and graphed. The locus answers the question, “for any given equity weight on the costs or benefits of the poor, what is the equity weight on the costs or benefits of the median necessary for the policy to breakeven in terms of net benefits (in other words, the breakeven equity weight on the median). For example, if a policy generates costs for the poor and the median, and benefits for the wealthy, we might ask, if we place a weight of 2 on the cost to the poor, what would the breakeven weight on the median be? Note that the locus also answers the question, “for any given weight on the median, what is the breakeven weight on the poor.” That is why we call it a locus of breakeven weights. Each point on the locus is a pair of weights such that the policy would breakeven.

Figure 1. Breakeven locus for a policy with costs to poor and median.

![Figure 1](image)

Figure 1 shows the breakeven locus for a policy which, after welfare weighting, generates costs of 5 to the poor and 10 to the median, and benefits to the wealthy of 30, generating a net benefit in terms of welfare efficiency of 15. It shows, for example, that if we equity weighted the costs to the poor, relative to the wealthy, by 2.5, we would need to apply an equity weight of 1.75 to the costs of the median for the policy to breakeven. If we raise the weight on the poor while holding the weight on the median constant or raise the weight on the median while holding the weight on the poor constant, the net benefit of the policy will become negative, and the status quo will be preferred. If we lower either of the weights, holding the other
constant, the policy will be preferred. This information allows decision-makers to consider whether the importance they place on the welfare of the poor and the median, relative to the wealthy, is sufficient to outweigh the aggregate welfare benefit of the policy. If a policy maker concludes that they consider the welfare of the poor to be worth at least 150% more than the welfare of the wealthy, and the welfare of the median to be worth at least 75% more than the welfare of the wealthy, they should stick with the status quo.

Note that the portion of the locus above the 45-degree line is dashed. In this region, the weight on the median is greater than the weight on the poor, which would violate intuition about the relative concern society should have for the two groups.

In figure 2, we consider a policy that has costs to the poor and benefits to both median and wealthy. In particular, suppose that after welfare weights, the cost to the poor is 35, the benefit to the median is 15, and the benefit to the wealthy is 40. Without equity weighting, the net benefit is 20. The locus of breakeven equity weights is upward sloping in this case, reflecting the fact that as the costs to the poor are inflated, the total net benefit falls, requiring the benefits to the median to be inflated in order to bring net benefits back up. Now the policy is preferred above the locus. For example, if the equity weight we place on the poor is 1.75 and the weight we place on the median is 1.5 (a point above the locus), the net benefit of the policy becomes positive. Starting with the same weight on the poor, if we instead place a weight on the median of only 1.25, which is just below the locus, the net benefit becomes negative.

Figure 2. Breakeven locus for a policy with costs to poor and benefits to median.

Again, we feel that the information contained in these graphs is intuitively meaningful and could be of use to decision-makers. However, for a time constrained decision-maker without a certain degree of familiarity with graphic presentation of numbers, the interpretation figures might be prohibitively complex.
<table>
<thead>
<tr>
<th></th>
<th>Study Data Sources</th>
<th>Elasticity Estimate</th>
<th>25\textsuperscript{th}</th>
<th>50\textsuperscript{th}</th>
<th>75\textsuperscript{th}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 1</strong></td>
<td><strong>Estimated Welfare Weights at Selected Income Percentiles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Convex Time Budget Experiments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andreoni &amp; Sprenger (2010)</td>
<td>97 U.S. subjects who made multiple choices among alternative budgets</td>
<td>.127$^1$</td>
<td>1.08</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>Acland (2022)</td>
<td>91 U.S. subjects who made multiple choices among alternative budgets</td>
<td>2.02</td>
<td>3.17</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Indirect behavioral evidence: Consumer Demand Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evans &amp; Sezer (2002)</td>
<td>1967-97 U.K. Government time series data on per capita income, food expenditures, and food prices</td>
<td>1.6</td>
<td>2.49</td>
<td>1.00</td>
<td>0.45</td>
</tr>
<tr>
<td>Demuynck (2018)</td>
<td>2007 U.S. Consumer Expenditure Survey data on food</td>
<td>0.3$^2$</td>
<td>1.19</td>
<td>1.00</td>
<td>0.86</td>
</tr>
<tr>
<td>Groom &amp; Maddison (2019)</td>
<td>1964-2010 U.K. Government time series data on food and non-food commodity expenditures</td>
<td>3.57$^3$</td>
<td>7.67</td>
<td>1.00</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Indirect behavioral evidence: Lifetime consumption model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groom &amp; Maddison (2019)</td>
<td>1975-2011 quarterly U.K. government data on non-durable goods and services, as well as time series on prices and the real interest rate.</td>
<td>1.58</td>
<td>2.46</td>
<td>1.00</td>
<td>0.46</td>
</tr>
<tr>
<td>Blundell, Browning, &amp; Meghir (1994)</td>
<td>1967-97 U.K. Family Expenditure Survey time series data on income and consumption of food, alcohol, fuel, clothing, transport, services and other goods</td>
<td>1.2 - 1.4$^4$</td>
<td>2.10</td>
<td>1.00</td>
<td>0.53</td>
</tr>
<tr>
<td>Havranek et al (2013)</td>
<td>Meta-analysis of estimates from 169 published studies covering the U.K. &amp; U.S., as well as 102 other countries</td>
<td>2.0 (U.K.$^5$)</td>
<td>3.13</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7 (U.S.)</td>
<td>2.64</td>
<td>1.00</td>
<td>0.43</td>
</tr>
</tbody>
</table>
The following formula is used to compute the weights: 

\[
\frac{y^m}{y^i} e
\]

where \( y^i \) is the income of the \( i \)th percentile group, \( y^m \) is the median income of the group at the 50th percentile, and \( e \) is the elasticity of the marginal utility of income.

Source for 2021 income at each percentile: U.S. Census Bureau’s Annual ASEC Survey in September 2021.
https://dqydj.com/average-median-top-household-income-percentiles/

---

1 Andreoni and Sprenger (2010) estimate the function’s curvature parameter, \( \alpha \), which has values between .706 and .977. In our view, the best estimate is .873, and that is the estimate we use in computing the welfare weights. Because \( 1 - \alpha = e \), the elasticity approximates 0.127.

2 The estimated income utility curve is non-linear, downward sloping, and irregular. Thus, it does not have a constant elasticity. Demuynck (2018) provides welfare weights in Table 4. Given these values, equation (3) can be used to determine the implied welfare weights at different points along the utility curve. For example, at the 25th percentile, it is .267; at the 75th percentile, it is .272; and at the 90th percentile, it is .325.

3 Confidence interval includes zero.

4 Weights are computed from the mid-point of this range.

5 The overall mean for all countries is 2.0 The reported means are all unweighted.

6 Elasticities for 30 additional countries were estimated with values between 0.81 and 4.99.

7 Layard, Nickell, & Meghir (2008) also used similar survey data for European countries to estimate the elasticity of the marginal utility of income for those countries. Almost all the resulting estimates were between 1.15 and 1.35.

8 The median value is 1.00. The standard deviation is 0.85 indicating that the responses varied considerably.

9 The elasticity falls to 1.28 if National Insurance contributions are incorporated into the tax rate measure.

10 Elasticities for four additional OCED countries were estimated with values between 1.3 and 1.6.

11 Elasticities for 18 additional OCED countries were estimated with values between 1.17 and 1.74.