Measuring the Costs of Climate Change Policies: Issues in the Interpretation of Common Metrics¹

W. David Montgomery

Anne E. Smith

Sharon L. Biggar

Charles River Associates

Abstract

Studies of the costs of climate change policies have utilized a variety of measures or metrics for summarizing costs. The leading economic models have utilized GNP, GDP, the "area under a marginal cost curve," the discounted present value of consumption, and a welfare measure taken directly from the utility function of the model's representative agent (the "Equivalent Variation"). Even when calculated using a single model, these metrics do not necessarily give similar magnitudes of costs or even rank policies consistently. This paper discusses in non-technical terms the economic concepts lying behind each concept, the theoretical basis for expecting each measure to provide a consistent ranking of policies, and the reasons why different measures provide different rankings. It identifies a method of calculating the "Equivalent Variation" as theoretically superior to the other cost metrics in ranking policies.

DOE Patent Clearance Granted

Mark P. Dvorscak

Mark P. Dvorscak

(630) 252-2393

E-mail: mark.dvorscak@ch.doe.gov

Office of Intellectual Property Law

DOE Chicago Operations Office

7.8.03

Date

¹ Research supported by Grant No. DE-FG02-00ER63035 from the U.S. Department of Energy. All statements in this paper are solely the responsibility of its authors.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

1 Introduction

The Kyoto Protocol signed in December 1997, established specific emission reduction targets for greenhouse gases that are to be achieved by 2010. While the agreement thus defines precise national objectives for reducing emissions, it leaves open the methodology by which such goals are to be reached. For this purpose the Protocol includes several alternative proposals, including provisions for international emissions trading and credits for reforestation, amongst other possible policies. The open-ended nature of the agreement has provided signatory countries with the opportunity to evaluate which climate change strategy provides the least-cost course of action to achieve the required emission reductions. It has also provoked continuing discussions of the "Cost of Kyoto" in many countries that are assessing the implications of these commitments.

To aid in the decision over the most suitable policy - for both individual nations and the global economy as a whole - economists have developed sophisticated models that simulate the implications of differing approaches. These models provide valuable input into the policy debate, by creating summary measures of the costs and benefits of climate change policies that allow differing strategies to be ranked by objective criteria. However, where it has been possible to compare the metrics used in each of these models, it has been noticeable that they produce different and sometimes contradictory results. This has caused some concern amongst policy-makers who rely on the quantitative estimates produced by these models.

The aim of our research for DOE has been to investigate one of the reasons why differing economic models can provide contradictory results - their adoption of different conventions for calculating and reporting costs -- and to assess the sources of bias and variability in the cost metrics utilised in such models. In all, the objectives of this research are three-fold:

1. To provide a theoretical foundation for determining the most appropriate cost measure for ranking different policies;
2. To provide a critical review of cost metrics in climate change models in current use; and
3. To undertake an empirical analysis, of why and to what extent current measures differ from the theoretical ideal.

This paper is one of two produced in the course of our research, and essentially addresses the objectives (1) and (2) above. It explains the theoretical foundations in welfare economics of an ideal economic impact metric, and explores its relationship to other commonly used cost metrics. This lays a foundation for an empirical analysis of the numerical differences among the alternative metrics of economic impact, and the determinants of those differences. A second paper provides that analysis. This paper is written in relatively non-technical terms and is targeted to an audience that wishes to understand better the various types of economic impact metrics that they will

encounter in reports from multiple greenhouse gas policy models. It also provides motivation for the empirical research recorded in the second paper. The second paper is written in a more technical style, and addresses some of the issues noted in this paper in greater mathematical detail.

This paper is structured as follows. Section 2 summarizes the theoretical foundations of an optimal economic impact measure. It begins with a description of the Pareto criterion that provides a guide for choosing the optimal policy to maximise the welfare of the individual agents within a society. We then show how (with the use of a money metric utility function) cardinal rankings of individual preferences for policies can be obtained, and the change in the welfare associated with a policy shift calculated. This leads us to the welfare measure theoretically preferred for policy analysis: equivalent variation (EV). EV is then compared to another welfare measure that is widely used in practice: consumer's surplus (CS).

In Section 3, we turn to a range of measures commonly used in estimating the impacts of a greenhouse gas policy, explaining how each is related to the theoretically ideal measure of EV. These measures include:

- Direct costs,
- GDP, and
- Present value of future potential consumption.

Section 4 provides numerical examples from CRA's MS-MRT model demonstrating the extent to which the various forms of impact estimates can produce different results, even when derived from the same basic model. We demonstrate that:

1. The magnitudes of the measures are not directly comparable in absolute or in relative terms – in particular, GDP, the Direct Cost and the area under the marginal cost curve are annual measures whereas PVC and EV are intertemporal measures;
2. The measures can produce different policy rankings; and
3. The measures can even be directly contradictory.

In the second paper, we turn to the question of what drives these different outcomes.

An appendix supplements the discussion of Section 2 with a more mathematical derivation of EV, and its comparison to CS.

2 Objective criteria for ranking policies in terms of how they affect social welfare

It is desirable to know whether or not a change in government policy will improve the aggregate welfare derived from an economy. The following discussion sets out a procedure for deciding whether or not a policy is beneficial overall. The methodology involves three steps:

1. A logical and objective method for identifying whether or not a policy will improve overall social welfare;
2. A procedure for quantifying the welfare impact on individuals; and
3. A method for aggregating welfare changes across all individuals within a society.

The most fundamental principles in welfare economics are the Pareto criterion, plus an extension to it known as the Compensation Criterion. These criteria guide us to objective measures for evaluating whether or not a policy would improve aggregate welfare, or "satisfaction". We first summarize these criteria, then consider how to quantify the changes in satisfaction from differing policies, to determine whether a policy meets these criteria. The Pareto and Compensation criteria suggest that societal welfare changes are estimated by aggregation of individual changes. Thus, our discussion of the theoretically correct approach to quantification of welfare changes starts with estimating welfare changes experienced by individuals. Our discussion then turns to the methodologies and difficulties associated with summing individual welfare measures into an overall social cost metric. Whether stated at the individual or aggregate level, the term used for this welfare improvement is "Equivalent Variation" (EV). The name is given because the underlying calculation is of the *variation* in income that is *equivalent*, in its effect on welfare, to the policy under evaluation.

The following discussion of policy impacts addresses only the cost side of policies. It includes both direct costs and indirect costs, but excludes the welfare implications of the policy's benefits, such as the value to consumers of improved ecosystem services that may result from reducing greenhouse gases. These additional welfare impacts are also important to include in any comprehensive cost-benefit analysis, but this paper confines its discussion to the theoretical and practical complications in assessing the economic impacts of achieving the policy constraints. Estimation of welfare-enhancing impacts of environmental benefits from such a program relies on the same welfare concepts, but the issues for their practical application are substantially different and are not discussed here. This limitation of the scope of the discussion does not, however, limit the usefulness of rank-ordering policies on the basis of costs only. As long as the costs being compared all lead to the same level of total greenhouse gas emissions in the same time frame, then the benefits, whatever they may be, would be identical from policy to policy. Thus, any rank-ordering on the basis of costs alone will be identical to the rank-ordering if benefits estimates were to be included. All of the comparisons of

policy impacts discussed here differ only in the flexibility of where emissions are reduced, and are all comparable in terms of the emissions constraints that they achieve.

2.1 PARETO CRITERION AND COMPENSATION CRITERION

How can a society choose between policies? When can we say that Policy A is 'better' than Policy B?² These are questions that have been addressed by welfare economics, and codified as the Pareto and Compensation Criteria. In the simplest terms, these two criteria say that Policy A is 'better' than Policy B if:

- (a) Everyone is better off under A (the Pareto Criterion), or
- (b) 'Winners' under A can compensate those who are 'losers' under A to the point that all are better off (the Compensation Criterion).³

The Pareto Criterion for selecting one possible future over another is a simple starting point that is difficult to reject. If everyone affected by a Policy A unanimously agrees that Policy A is superior for themselves individually, then who would ever argue against choosing that policy?

Such a simple choice rarely presents itself, however. In actuality it is probable that some individuals would prefer Policy A, while others would prefer Policy B. The Compensation Criterion was developed to address this more complex but more common situation. The Compensation Criterion recognizes that as long as winners under Policy A could give some of their gains to the losers, and compensate them for all their losses, without becoming losers themselves, then Policy A, combined with some side transactions, could be transformed into a policy that meets the Pareto Criterion – Policy A provides a *potential* Pareto improvement over Policy B.

Note that Compensation Criterion does not require that any compensation actually occur; reallocation of gains does not need to be built into the policy in order for Policy A to be deemed better than Policy B under the Compensation Criterion. There are two ways to interpret this view. Philosophically, the Compensation Criterion reflects a utilitarian world view, that the welfare of a society is the sum of the well-being of all of its individuals. This is no longer a popular political philosophy. More pragmatically, proponents of Policy A should have personal incentives to offer some form of compensation to the losers, because this will enable adoption of a policy that would enhance their individual level of well-being. Thus, Policy A should be possible to pass through democratic political processes which allow for "horse-trading" and "log-rolling", making it a reasonable one for policy makers to support. This idea led at least once in the development of this concept to the thought-experiment of a Department of

2 In this discussion, one could think of Policies A and B as two alternative ways of achieving a given change, such as a given emissions target. Alternatively, Policy A could be viewed as a new policy, and Policy B the status quo without that policy (i.e., a "Base case")

3 'Winners' are defined as individuals who have a higher level of satisfaction as a result of the policy implementation.

Efficiency and a Department of Distribution. The Department of Efficiency is tasked to apply the compensation criterion, and the Department of Distribution is tasked to ensure that the resulting gains are redistributed in a socially desirable way.⁴

With these criteria, founded in the individualist, democratic precepts that characterize our society, we have the grounds for analyzing and comparing among policies in a quantitative manner. In order to evaluate whether or not under the appropriate reallocation of costs, individual "winners" who prefer A to B could compensate all 'losers' who prefer B to A, it is necessary to be able to explicitly measure the change in welfare of all individuals in society, under differing policies.

2.2 WELFARE, UTILITY, AND EQUIVALENT VARIATION

In welfare economics, welfare improvements are represented conceptually as increases in individual "utility". Utility is the satisfaction that an individual derives from consuming goods, services and leisure.⁵ Utility is not simply the sum of the monetary value of all things consumed because individuals derive different degrees of satisfaction from different types of goods. Further, increasing amounts of most goods produce a decreasing rate of increase in an individual's degree of satisfaction. Utility is therefore a *function* of consumption; the function applies different weights to different types of goods, and the utility function usually starts to flatten out as consumption of each good is increased.

Economists do not assign any absolute interpretation to the utility function. Rather, the theories of consumer choice and behavior that have been built on the utility function concept rely only on the *relative* ranking that such a function implies for consumption of alternative "bundles of goods and leisure." Thus, there is no single "correct" utility function, because there will be an infinite set of utility functions that all can equally well explain the same set of choices made by an individual.⁶ Among the many possible utility functions that can "explain" consumer behavior, it is possible to calibrate one to a monetary unit such as current dollars. When this has been done, it is possible to interpret the welfare impact projected for a policy in terms of its effective impact to current spending power of consumers. This provides a theoretically coherent estimate of the "cost" of that policy to consumers.⁷

⁴ Richard A. Musgrave, *Principles of Public Finance*. It should be noted that the compensation criterion can produce contradictory results, as pointed out by Scitovsky.

⁵ In the rest of this paper, we will use the term "goods" to refer to all goods and services generally. Leisure will be kept as a separate concept, since some models account for it and others do not.

⁶ For example, taking one utility function that explains my economic behavior and multiplying it by 2 will provide exactly the same ranking of "Bundles of goods and leisure", though each "bundle" will nominally provide twice the utility. Technically, any function that is a monotonic transformation of a proposed utility function is considered an equally valid utility function.

⁷ This type of utility function is known as a "money metric utility function", and is described mathematically in the appendix.

The maximum level of satisfaction achievable at any point in time depends on one's income and also on prevailing prices. The economic impacts of a policy are felt by consumers when the policy changes the prices and/or incomes that they face. The monetary measure of the change in an individual's utility due to a policy is equal to the specific amount by which the individual's income would have to be adjusted, after accounting for the policy's anticipated price and income impacts, in order for that individual to be able to achieve a level of utility *equivalent* to today's level. This measure is known as "Equivalent Variation" (EV). Being linked to income, it is stated in monetary terms. By construction, it is measured in current dollars and therefore relatively easy to interpret.⁸

EV may not be a widely recognized term among laymen, but it is the theoretically correct monetized measure of a policy's impacts to consumer welfare. Conceptually, it tells us an individual's "willingness to pay" to have a policy implemented, or to avoid having a policy implemented (depending on whether the policy will affect the individual positively or negatively). The appendix to this paper provides a mathematical formulation and derivation of EV, and one might also refer to the textbook by Varian.⁹

One reason EV has not been a common fixture in policy analyses is the difficulty in calculating the measure in models that do not start their representation of behavior with an individual utility function. Direct estimation of EV requires that a model contain a direct representation of an intertemporal utility function. "Dynamic general equilibrium" (DGE) models have this feature, and this is one of the reasons for their growing importance in climate policy cost analysis.

However, greenhouse policy cost analyses are still frequently performed using models that do not have a capability to directly estimate EV. Most policy analysis models are constructed around more easily estimated demand curves, and/or are focused on technology processes that provide cost estimates, but not welfare estimates. In principle, EV can instead be estimated by "recovering" the underlying utility function implied by a demand curve that is posited in a model.¹⁰ In some models, the demand curves may be such that the underlying utility function could be recovered, but the calculation is complex and tedious. In many cases, however, the model's demand relationships were specified without concern for whether they would be computationally tractable for estimation of EV, and the necessary calculation is impossible. Although the latter types of models can produce measures of direct cost,

⁸ There is a similar concept to EV called Compensating Variation. It has a parallel definition except that the necessary income adjustment (i.e., the monetary value of the utility change) is stated in terms of the prices that would hold after the new policy has been brought into force. Since these altered prices are the subject of uncertainty, Compensating Variation is not viewed as a practical alternative to EV. It has primarily theoretical interest, and therefore is not discussed here. The technical appendix does describe it, however.

⁹ H. Varian, *Microeconomic Analysis*, 3rd edition, New York, N.Y.: W. W. Norton & Company, 1992. (See especially Chapters 7 and 10.)

¹⁰ "Recovering" the utility function requires finding the mathematical integral of the demand curve. Not all functions that may be used for a demand relationship can be integrated.

or macroeconomic metrics such as gross national product (GNP), none of them can produce the ideal measure of EV. Hence, it is important to understand how these various other measures of "economic impact" compare to EV, which we turn to in Sections 3 and 4.

2.3 RELATIONSHIP BETWEEN EQUIVALENT VARIATION AND CONSUMER'S SURPLUS

Before turning to the relationships between EV and other impact measures that are often produced by greenhouse gas policy analysis models, it is useful to note the relationship of EV to the concept of "consumer's surplus". Consumer's surplus can be loosely defined as the total difference between what buyers are (hypothetically) willing to pay and the price that they actually pay in the marketplace. Changes in consumer's surplus are often used to estimate welfare losses associated with changing prices.

Figure 1 illustrates the definition of consumer's surplus. A standard demand curve is estimated from observable market place data, as drawn in Figure 1 for a hypothetical "Good 1". If the market price of Good 1 is P^0 , and a quantity Q^0 is sold at that price, the consumer's surplus associated with Good 1 can be represented as the area to the left of the demand curve and above the price (the grey triangular area).

Changes in consumer's surplus are based on how the area of this triangle are altered by altered prices (or other changes to the market outcome). If the price of Good 1 were to be higher than it is in Figure 1, then there will be less consumer's surplus. Figure 2 shows the loss of consumer's surplus that would be estimated for a price increase from P^0 to P^1 – it shows both a reduction in consumption of Good 1 (the triangular area from Q^1 to Q^0), and, for that part of consumption of Good 1 that remains, a reduced utility because a greater amount must be spent to obtain the utility of consuming Good 1 (the remaining rectangular portion of the shaded area to the left of Q^1).

Figure 1. Consumer's surplus

Figure 2. Change in consumer's surplus due to a price increase

In Figure 2, the change in consumer's surplus is calculated from the point of view of a single consumer with a fixed income. In a full model of the economy, there many consumers; in aggregate, consumers own all resources and production facilities. Consumers therefore in aggregate also receive all the additional income that comes from selling goods at higher prices.¹¹ This means that in a model of the full economy, the portion of consumer's surplus attributable to paying higher prices for that amount of Good 1 that consumers continue to purchase (i.e., the rectangular portion to the left of Q_1) flows back to consumers in the form of increased income. This means that the remaining triangle is the only loss to aggregate consumer's surplus in an economy-wide model. This net economy-wide loss in consumer's surplus is called the "deadweight loss" of a policy that increases the price of a good. It is sometimes also referred to as the "Harberger Triangle."

Thus, with data to estimate a demand curve, one can estimate the loss in consumer's surplus due to price changes. This sounds much like the concept of EV, and there is a relationship. However, EV remains the more exact measure. The limitation of consumer's surplus is that it makes no provision for possible losses of spending power, or "income effects" that may occur when prices rise. If an individual's income is fixed, and Good 1 represents a large portion of the household budget, then a rise in the price of Good 1 will cause total spending power to fall. This reduction in spending power may cause consumption of Good 1 to fall by more than would be predicted by the standard demand curve that has been drawn as a function of the price of Good 1 only. EV accounts for both income effects and price effects, whereas the traditional method of computing consumer's surplus does not. As a result, EV losses due to a price

¹¹ Consumers in aggregate also ultimately consume or receive as income all tax revenues.

increase ought to be larger than the loss estimated by the consumer's surplus approach.

Without going into the mathematical proof here,¹² the loss in EV due to a price increase for a particular good is always greater than the loss in consumer's surplus, except under one special set of circumstances, where they are actually identical in magnitude. They are identical when the demand for the good in question depends only on price; i.e., that as income increases the demand for the specified good does not increase. Examples of goods where this may occur are those which represent a very small part of the total household budgets, such as pencils or potatoes might be in an average U.S. household. Goods that are not likely to hold this special property include heating, air conditioning, transportation, and housing – key goods of interest for greenhouse gas policy impacts.

Thus, in general, EV is closely related to consumer's surplus, but EV will be the better measure of economic impact for greenhouse gas policy costs. In most instances consumer's surplus will provide a biased measure of welfare impacts, and when it is not biased, it is still no better than EV. This relationship is important to know because, as will be shown in Section 3, many of the forms of policy cost estimates not based on EV are usually justified because they are approximations of consumer's surplus, rather than approximations of EV.

2.4 AGGREGATION

The last step necessary for an objective ranking of policies is to aggregate the estimates of EV of the many affected individual consumers. DGE models tend to use just one or a few "representative" consumers to reflect the wide variation of types of people. This is due to lack of data at a more disaggregated level on which to base the model, and also for computational reasons. An important technical question is whether EV estimated for a "representative" consumer provides an unbiased estimation of the EV that would be obtained if we were to directly estimate and sum the EVs for all individuals.

One can show that a certain class of utility functions will provide an unbiased estimate of the aggregate welfare when estimating EV from a single consumer that represents an entire group of individuals. The form of function known as the constant elasticity of substitution (CES) is one of the acceptable forms for this purpose. This fact plus several other computational advantages of CES functions cause them to be widely used to directly represent the utility function in DGE models.

Although this aggregation issue is of a more technical interest than the goals of this paper, we note it here for the sake of completeness. We also note that the CES form of utility function does *not* produce demand functions for goods that are independent of

¹² For a proof, see Varian, op. cit., pp. 167-168.

income effects.¹³ While this is appropriate and desirable for the application of interest, it means that DGE-based estimates of EV will *not* be expected to coincide with standard estimates of consumer's surplus.

2.5 DEALING WITH INTERTEMPORAL TRADEOFFS

The long time scales for climate change, and the tradeoffs between near term costs and future benefits, make the treatment of intertemporal choice paramount. This is not simply an issue of finding the right discount rate to calculate the present value of consumption (though that is likely the most important practical problem), but includes issues of how the utility function treats tradeoffs over time (it need not be equivalent to the PV of the instantaneous consumption aggregate at a fixed discount rate). There are also modelling issues tied up in the choice of welfare measure, especially for the choice of time horizon.

3 Equivalent variation compared to other cost measures

Section 2 has introduced the concept of equivalent variation (EV), explaining how it is the ideal measure of welfare change. However, not all models in current use are able to estimate EV. In this section we review the following alternative measures of economic impact of climate change policy that are in circulation:

- Discounted Present Value of Consumption (PVC)
- Direct Cost (Area under Marginal Cost Curve)
- Gross Domestic Product (GDP)

Table 1 indicates the measures used in different models participating in the Energy Modeling Forum. We explore their bases in economic theory, and likely relationship to EV. We also describe some of the limitations associated with each. Table 2 summarizes the key limitations that we will highlight.

The type of measure used to assess a policy's economic impact is tied to the fundamental structure of a model. Only models that have a microeconomic foundation of utility and profit maximization can estimate EV. DGE models with representative utility-maximizing agents are the only ones that can compute EV directly. Models that incorporate supply and demand relationships that have been derived from utility and

¹³ Mathematically, an unbiased aggregate utility estimate requires a utility function in the Gorman form. CES functions and quasilinear utility functions both satisfy this condition, but the CES function is not quasilinear. However, only quasilinear utility functions can produce the conditions where EV and consumer's surplus coincide.

profit maximization can be used to calculate consumer surplus measures or, under certain conditions, EV. However, many macroeconomic models do not have any microeconomic foundation. These models are limited to calculating either the discounted present value of consumption or GDP-type measures.

Table 1. Summary of metrics used by leading greenhouse gas policy models

Model	Time Horizon	Primary Cost Metric	Other Metrics Utilized	Saving and Investment
MERGE	2100	GDP	PVC	Dynamic Utility Maximization
SGM	2020	Area under MAC curve		Fixed
EPPA	2100	PVC	GDP	Myopic
RICE	2300	PVC		Dynamic Utility Maximization
MS-MRT	2030	EV	All except GNP	Dynamic Utility Maximization
ABARE	2010	GNP	Consumption	Saving Function
G-Cubed	2020	GNP	GDP	Saving Function

Table 2. Summary of limitations and difficulties with impact metrics

Impact Metric	Limitations or Difficulties
Equivalent Variation (EV)	<ul style="list-style-type: none"> Imprecision due to approximation of infinite time horizon. Possibility of using inconsistent discount rates Assumes constant marginal utility of income Requires highly restrictive conditions to be a measure of welfare Leaves out terms of trade effects Annual measure Price index issues Distorted by use of gross, not net investment Oversimplifies relationship between consumption and investment Ignores changes in foreign indebtedness Same as GDP, except accounts for payments and receipts from abroad
Present Value of Consumption (PVC)	
Direct Cost	
Real GDP	
Real GNP	

While we will focus in this paper on the comparison of the ideal of EV to present value of consumption, direct costs, GDP, and GNP, it is also important to note that even within the class of models that can estimate EV, there may be serious complications when addressing intertemporal problems such as climate change poses.

- Some of the relatively simple models do not actually perform an intertemporal utility maximization to assess changes to investment decisions. They may instead apply a static utility maximization recursively, or some other ad hoc rule for determining saving and investment. These models prevent a logically consistent computation of EV despite having microeconomic foundations.
- Models with the requisite intertemporal utility maximization (i.e., *dynamic* models) for calculation of EV face thorny problems of dealing with estimation of impacts over an infinite time period. That is, at some point the model calculations must be truncated, and the “rest of time” approximated. The key is to truncate the model at a point where the economy has returned to a new steady state after the initial “shocks” caused by the policy. Nevertheless, imprecision associated with the “rest of time” approximation can appear in the EV estimated by intertemporal models. This is as a concern to be alert to even with the most sophisticated DGE models, and hence is noted in Table 2.

3.1 PRESENT VALUE OF CONSUMPTION (PVC)

In Section 2, we explained that utility is fundamentally tied to consumption, although not exactly equal to the sum of the value of goods consumed (because of differing preferences for different goods and diminishing marginal utility from consuming more and more of any one good). Further, a dynamic equilibrium is one where the present value of the utility of consumption is maximized, suggesting that higher EV will be associated with higher consumption levels. This fundamental relationship suggests that there should be correlation between consumption levels and EV levels from one policy to another. Almost any model type computes estimates of consumption in each time period, and so the question arises whether the present value of consumption (PVC) projected by a model might be a useful proxy for EV, at least for the purpose of creating a relative ranking of policies.

Indeed, several well-known greenhouse gas policy models report PVC as if it were the equivalent of EV. For example, the EPPA model of MIT reports PVC changes under the label “change in welfare.” Reliance on PVC is necessitated in this case because EPPA uses a recursive rather than dynamic formulation, and does not include an intertemporal utility function. PVC is also the primary welfare-oriented impact metric reported from the MERGE and RICE models, though both these models are constructed so that in principle they could estimate EV.

The present value of consumption is defined as the net present value of aggregate consumption over the time frame (or horizon) of the model (Equation 1). Consumption consists of private spending on goods and services (such as food, housing, and clothing) and public spending on services (such as health, education and welfare). In order to rank policies that affect the welfare of individuals in the future

the measure incorporates a rate of time preference that re-values future consumption in today's present value terms. In other words a weighting (or discount rate) is applied to future consumption in order to be able to compare it with the value of current consumption today.

Equation 1. Present value of consumption

$$PVC = \sum_{t=1}^T \frac{C(t)}{(1+w)^{t-1}}$$

where:

C(t) is the aggregate consumption in period t; and
W is the social discount rate applied to future consumption.

As with EV, a key problem with using PVC as an aggregate welfare measure stems from the difficulties of measuring economic well-being over an infinite horizon. It is not yet possible to undertake computations that run to infinity. For this reason so-called 'infinite horizon models' are, in implementation, terminated at some finite point. Problems with measurement arise in determining how to account for the effects of climate change policies that occur after this specified termination cut-off. In most instances this is dealt with by making an adjustment to consumption that accounts for the change in the value of the capital stocks estimated to exist at the point in time that model calculations are cut off (i.e., the "terminal capital stock"). In climate change models, in particular, it is important that the valuation of terminal capital stock is accurate, as emission limits are known to produce large drops in investment (as high as 2.5 per cent under a "no trading in the US" scenario).¹⁴ These changes in investment over the shock period can imply very different terminal capital stocks, and hence very different long-term welfare outcomes that need to be accounted for.

Another possible discrepancy between PVC and the underlying welfare measure arises if the chosen intertemporal utility function does not imply a constant discount rate on consumption (or if a different consumption discount rate than that implied by the utility function is used). In this case, PVC will not fully account for preferences regarding the distribution of changes in consumption over time. It is also possible that the discount rate chosen for calculating PVC will not be consistent with the underlying intertemporal utility function. In principle, this discrepancy can make policy changes that reduce welfare measured by the value of the underlying utility function appear to increase the PVC, and vice versa. However, for standard utility functions and the

¹⁴ Investment falls because consumers buffer the immediate and largest potential impact on their consumption levels during the near-term periods by reducing investment during that time.

magnitude of impacts of policy changes investigated, this does not appear to be a large problem in practice.

Thus, while there are some potential differences, PVC may serve as a reasonable estimate for a welfare impact. In many cases, it is already used in that manner. It is, nevertheless, not exactly identical to EV.

3.2 DIRECT COST

The PVC and EV are both firmly rooted in welfare economics. Many energy models, however, are more oriented to making direct estimates of the costs of achieving a particular policy objective. This is particularly true of "bottom up" models that represent energy technologies and their substitutes in specific detail. (DGE models, in contrast, use continuously differentiable production functions that often leave the specific technology choices quite abstract.) The SGM model uses direct cost estimates rather than direct welfare impact estimates as its primary impact metric.

Although direct cost estimates may not be derived from a utility model, they do in fact have a link to the concept of consumer's surplus. This is not a direct link to EV, but as discussed in Section 2, it may serve as a rough approximation.

Area under the marginal cost curve

The usual method for estimating direct costs of achieving a specific emissions reduction is to estimate the marginal cost curve and then calculate the area under that curve up to the reduction target.¹⁵ The relevant marginal cost curve can be constructed relatively simply from models that can address emissions trading or emissions taxes. The procedure would be to run the model in question to produce estimates of the carbon price associated with different caps on carbon emissions, and to plot the price of carbon against the emission reduction.

A simpler approximation is actually more commonly applied. Rather than plot out the entire marginal cost curve, the analyst will simply obtain an estimate of the marginal cost *at the point of the target emissions reduction*. This is the same as the projected emissions permit price. The estimated total cost for that reduction target is calculated as the $\frac{1}{2}$ the marginal cost times the amount of emissions reduced, based on the assumption that with zero emission reduction marginal cost is also zero. Figure 3 illustrates how it is that this represents an approximation of the true total cost. The degree of error in this approximation is greater when the marginal cost curve is more curved over the range of approximation.

¹⁵ Mathematically, total cost is equal to the area under the marginal cost curve.

Figure 3. Approximation of total cost

Connection to welfare: Harberger Triangles

The preceding approximation is closely tied to a measure of the welfare impact of a policy known as the “deadweight loss”. Deadweight loss is a term used in tax analysis to refer to the loss in economic welfare due to a tax, over and above the actual revenue collected by the tax. For example, a sales tax on restaurant meals causes consumption of restaurant meals to fall. The tax causes a price increase, and therefore reduced consumption of restaurant meals. The loss in consumer surplus equals the revenue collected by the tax, plus an additional triangle representing the amount representing the loss in welfare from the meals foregone (see Figure ...) This remaining triangle in the area representing consumer surplus is the deadweight loss.

A perfect tax would collect revenue without causing any change in behavior, and thus would have no deadweight loss. The example most frequently used to illustrate a tax with no deadweight loss is a poll tax, in which each resident is charged the same lump-sum tax. (The example itself illustrates the problem of finding a tax with no distortions, because the poll tax can discourage having children, and encourage migration.)

The deadweight loss from a tax can also be approximated as $\frac{1}{2}$ the tax rate times the change in quantity purchased as a result of the tax. Returning to the demand curve of Figures 1 and 2, we now label the tax revenue collected on this good. This revenue is not a loss to the economy as a whole, and therefore does not reduce welfare. When a tax is imposed that raises prices, some part of the loss in consumer’s surplus (the rectangular part in Figure 2) is recycled through the economy when taxes revenues are spent by the government: there is no net loss to the economy as a whole. However,

the policy also has a “deadweight loss” where the loss to consumers does not reappear as gains to someone else. This is the portion of the response to higher prices that reduces total demand from Q_0 to Q_1 . This net welfare loss is the triangular area shaded in Figure 4, sometimes called the “Harberger Triangle”.

Figure 4. Harberger Triangle

If carbon emissions were proportional to the consumption of a single good, then the area under the marginal cost curve would be equal to the loss in consumer surplus from a policy that limited emissions. We can think of that good as an aggregate energy good, made up of different types of energy that must be consumed in fixed proportions. In this case, the only option for reducing emissions would be the reduction in consumption of the aggregate energy good, and the welfare loss from this reduction in consumption would be approximated by the Harberger Triangle. The price of a permit would be the price increase for energy required to reduce energy consumption by the required amount.

If Figure 4 is to be interpreted as a climate policy, the increase in the price of energy will equal to the price of carbon times the quantity of carbon in each unit of energy. Similarly, the reduction in quantity demanded of the good will be equal to the quantity of carbon per unit times the amount that carbon emissions are to be reduced. Thus, the area of the Harberger Triangle is equal to $\frac{1}{2}$ times the price of carbon times the quantity of carbon reduced times the good's carbon intensity. Treating the carbon intensity as a given, the size of the Harberger Triangle is directly proportional to the simple approximation of the area under the marginal abatement cost curve. Thus, any policy ranking based on the area under the marginal abatement cost curve will likely reflect a ranking based on direct calculation of Harberger Triangles, which is a measure of change in consumer surplus.

The Harberger Triangle will fail to approximate EV under exactly the same conditions that consumer's surplus fails to approximate EV: a good whose demand depends on the price of other goods and income. As noted in Section 2, the goods whose prices are

most likely to be affected by a greenhouse gas reduction policy are not likely to meet this condition.

None of this discussion of direct cost has mentioned what is the primary purpose of many energy models: examining energy supply, power generation and (possibly) technologies for producing other goods to determine the cost of replacing types of energy now in use with lower carbon forms of energy or with other resources. These costs indeed appear in the direct cost measure, and they also appear in the calculation of consumers surplus. Carbon limits or permit prices affect cost conditions in the economy. How these changes will affect the representative consumer depend on how they affect the relative prices of different goods consumed and also how they affect the income of the representative consumer. All profits earned by businesses in the economy are treated as income of the representative consumer, as is income from labor and income from capital investments. Determining how these changes in income and prices affect economic welfare requires a calculation of deadweight loss, and the direct cost measure of area under a marginal cost curve for carbon provides an approximation of this measure.

Limitations of direct cost

The key problems associated with using direct costs as a measure, are:

- Direct cost is an annual measure and cannot therefore, fully take account of the intertemporal nature of climate change issues.
- It relies on the change in consumer as an accurate measure of the overall change in welfare. As we have explained in Section 2, consumer's surplus (and for the same reasons producer's surplus) will only provide an accurate representation of welfare loss (or gain) under certain specific conditions.
- Estimation of Direct Cost from a single marginal cost estimate is further prone to error because it may be a poor approximation of the true total cost if marginal costs are highly nonlinear.

Additionally, many times direct costs are the primary focus in energy sector models. Pure energy-sector models often assume no economic adjustments outside energy sectors. One example is the use of a technology-based model to calculate the gross increase in cost of satisfying a given energy service demand, under a constraint on carbon emissions. These are inappropriate tools for assessing the impacts of a policy that will have wide-ranging implications throughout the entire economy, such as greenhouse gas policy. Although this is not a criticism of direct cost, per se, the need to rely on direct cost suggests that there may be a more fundamental inappropriateness in the choice of model.

3.3 GROSS DOMESTIC PRODUCT (GDP) AND GROSS NATIONAL PRODUCT (GNP)

Gross Domestic Product (GDP) is defined as the total value of a nation's output that is produced within its borders in a given period of time. Gross National Product (GNP) is defined as the total value of a nation's output that is *owned* by residents of that nation.¹⁶ GNP is arguably the most common measure of economic impact that is currently reported and it is widely used by policymakers, economists, international agencies and the media as the primary scorecard of a nation's economic health and well-being. It is the primary output of the G-Cubed and ABARE greenhouse gas policy models. GDP is also reported by MERGE, EPPA, G-Cubed, and MS-MRT. For this reason it warrants detailed attention in this analysis.

Although widely used, GNP and GDP changes are perhaps the furthest removed from a formal measure of welfare founded on microeconomics. It is defined as the sum in each calendar year of consumption plus investment plus government spending. Welfare associated with any particular year would be based only on the consumption measure. Investment is made in order to assure sustained or even higher future consumption, but it does not provide welfare in its own right. Thus, these widely used measures would not be expected to be inherently sound indicators of the welfare impacts of a policy.

Additionally, GDP and GNP change over time when the output of goods and services change and/or when the prices of those goods and services change. In other words, the nominal GDP measure incorporates movements in prices and it is possible that the measure can increase as a result of a general price rise, even if this price movement is not associated with an increase in output. In order to net out the effect of price changes Real GDP - which is calculated by dividing nominal GDP by the relevant price index - is often used as the preferred output measure. However, it is not easy to formulate theoretically correct price indices, and so even this correction can cause further inaccuracies.

While real GDP/GNP are thus superior in policy impact assessment to nominal GDP, there remain three main problems with using either of them to rank climate change policies. These are:

1. It is an annual measure;
2. It is associated with price index and numeraire problems; and
3. It is a measure of gross rather than net income.

We explain each in further detail below.

¹⁶ Output in GNP may not be physically located in the nation, but because of its ownership, contributes to the wealth of the nation where the owners reside. Conversely, GDP accounts for output based on where it is physically located, regardless of whether the ownership rights to the wealth are elsewhere. GNP is the better measure for a national welfare change, but GDP is a better measure of activity that affects a nation's employment levels.

Annual measure

Due to the lag between the time when emissions of greenhouse gases occur and the appearance of any possible effects that the accumulation of these pollutants may have on the global climate, the true costs of current emissions may not be felt for some time yet to come. Policies that attempt to limit greenhouse gas emissions are thus by their very nature intertemporal while GDP, is measured annually. For this reason GDP comparisons may not aid decisions that seek to rank climate change policies which are known to affect both current and future generations. For example, it is not uncommon for economic models to find that the change in GDP is positive in one period and negative in the next. In such an instance it is not possible to conclude that the policy is unequivocally superior to an alternative course of action, where that action may lead to a negative GDP change in this period but a positive GDP change the period following.

GDP may also be increased over a period of time by shifting resources from consumption to investment, or vice versa. Depending on whether resource allocation is optimal in the baseline, this shift may increase or decrease underlying economic welfare. This is seen in two kinds of results. One is a "buffering" of consumption during an economic downturn, by means reduced saving and investment, which can maintain GDP and consumption for some period of time, but at the expense of a smaller capital stock which eventually leads to lower productive capacity and output in the future. The other is a result of some policy stimulus, such as an investment tax credit, which boosts investment, possibly at the expense of consumption, and translates into a larger capital stock at some future point in time. Evaluating overall welfare effects requires including both the changes in consumption during the models time horizon with changes the capital stock at the end of the model horizon in an overall measure to determine whether the shifts between consumption and investment increase or decrease intertemporal welfare.

In short, the GDP measure provides no guidance as to the appropriate weightings (or discount rates) to apply to future changes in GDP and can only provide annual, and thus temporal, results for an essentially intertemporal problem.

Price index problems

As outlined above, movements in the general price level may influence nominal GDP changes. For this reason, Real GDP is the preferred measure as it removes the effects of inflation. However, in order to net out price movements, nominal GDP must be divided by a relevant price index, the choice of which is inherently problematic. In particular, models that do not include an explicit utility function must use a constructed price index, which implies a specific evaluation across commodities and time. In models that do include a utility function the ideal price index can be taken directly from the temporal expenditure function, but in these instances the results are very sensitive to the choice of numeraire.

In simulations performed with MRT, GDP reversals appear most frequently when policies that produce significant changes in capital flows are investigated, or when

there are large changes in international prices for non-energy goods. GDP reversals due to changes in capital flows occur because of shifts in investment as comparative advantage changes due to shifts in energy costs. In addition, the changes in international capital flows complicate welfare comparisons because the change in net foreign indebtedness must also be taken into account. A surge of investment in a developing country financed by foreign investment will increase GDP in the short and long run, but that portion of the capital stock will now belong to agents outside the economy in question – who will have claims on output of the economy. Capital inflows imply that an increasing amount of the productive assets of an economy are owned by agents located outside the economy. The resulting obligations to service the debt come out of resources that could otherwise provide consumption to residents of the economy in question. GDP measures the value of goods and services produced in the economy – the level of economic activity – not the income of residents of the economy, and therefore can give a distorted picture of resident welfare in the presence of capital flows. Use of GNP measures, which do account for repatriated profits, earnings of foreign workers, and debt service, should be investigated to address this issue.

Valuation of imports and exports provides a thornier problem, if the intention is to mimic welfare changes. The ideal measure would incorporate in GDP terms of trade gains or losses that occur because of changes in prices of imports and exports, but this is literally impossible if “Real GDP” is defined by fixed price weights. With fixed price weights, the same prices are used to evaluate Real GDP in the baseline and policy cases, and in every year, no matter how a policy changes those prices.

In some models, with endogenous imports and exports, changes in terms of trade will cause changes in quantities of imports and exports, and these changes will register in measures of “Real GDP.” However, the choice of price weights will affect the magnitude and the sign of the contribution of these changes in quantities of imports and exports on GDP.

These measurement issues are related to the manner in which changes in terms of trade result in different amounts of goods being available for final demand in the economy in question. For, if the price of imported oil falls, but all other prices remain the same, a smaller amount of other goods will need to be exported (or a larger amount of goods can be imported) to produce balance of payments equilibrium. The reduction in exports, all else constant, increases goods available for domestic use, and therefore welfare. Capturing these changes in real GDP requires some subsidiary calculations, either of the correct price index or a fundamentally similar calculation of changes in quantities. Models that calculate imports and exports endogenously will return to balance of payments equilibrium, by increasing some imports and decreasing some exports. This should lead to an increase in Real GDP whenever the price of an imported good falls. Whether Real GDP is evaluated at base period prices (before the fall in the price of imports) or current period prices (after the fall in the price of imports) will affect the magnitude of the calculated change. The standard Paasche-Laspeyres comparison will then apply – if quantity increases for a good whose price has fallen, and quantity decreases for a good whose price has increased, the increase in

Real GDP will be smaller than if quantity increases for a good whose price increases and decreases for a good whose price decreases.

Gross versus net income

GDP and GNP are measures of gross income, in the sense that they do not take account of the depreciation of capital. When policies lead to large reductions in investment for a sustained period of time, they lead to lower levels of capital stock and to reduced capital consumption allowances (CCA). As a result, GDP/GNP fall because of both contemporaneous reductions in investment and lower CCA. This effectively double-counts the reduction in investment and produces a larger percentage reduction in GDP than in net national income. GDP also fails to account for net interest payments on foreign debt, so that if a policy causes a significant change in foreign debt, GDP will inaccurately measure the change in real income for residents of a country. GNP theoretically corrects for this problem.

4 Numerical evidence of problems with alternative metrics

CRA's MS-MRT model is unusual in that it is capable of producing all of the above welfare measures at the same time, except for GNP. Thus, we can use MS-MRT to compare all of these measures against each other without being hampered by confounding due to differences in the functioning, parameters, and data across models. When such research has been undertaken three critical issues have come to light. These are:

4. That the magnitudes of the measures are not directly comparable in absolute or in relative terms – in particular, GDP, the Direct Cost and the area under the marginal cost curve are annual measures whereas PVC and EV are intertemporal measures;
5. That the measures can produce different policy rankings; and
6. That the measures can be directly contradictory.

Table 3 reproduces results from CRA's paper entitled, *"Effects of Restrictions on International Permit Trading: The MS-MRT model"*. Simulations of the Kyoto Protocol were run under the following policy scenarios:

- No emissions trading;
- Emissions trading under Annex B countries (both with Russian "hot air" trading and without); and
- Full global trading.

The table show each of the alternative metrics as produced by MS-MRT. Results are disaggregated by model region. They demonstrate the misleading and contradictory results that can arise when other metrics are used as a proxy for the theoretically correct measure of EV.

Table 3. Comparison of four welfare measures for selected regions

Scenarios	EV	PVC	Direct Cost		GDP	
			2010	2020	2010	2020
	(%)	(%)	(Billions of US\$)		(%)	(%)
<u>Scenario 1: There is no emissions trading (ET)</u>						
USA	-0.56	-0.41	-45.20	-43.55	-1.88	-2.33
Other Asia	-0.18	-0.03	n.a.	n.a.	0.51	0.01
China and India	0.34	0.59	n.a.	n.a.	0.64	0.21
Mexico and OPEC	-1.39	-1.65	n.a.	n.a.	-0.99	-1.60
Rest of World	-0.10	0.10	n.a.	n.a.	0.71	0.17
<u>Scenario 2: Annex B countries are permitted to trade</u>						
USA	-0.36	-0.31	-36.39	-57.53	-0.91	-1.27
Other Asia	-0.04	0.04	n.a.	n.a.	0.47	0.50
China and India	0.22	0.33	n.a.	n.a.	0.53	0.60
Mexico and OPEC	-1.15	-1.45	n.a.	n.a.	-0.64	-0.91
Rest of World	-0.08	0.01	n.a.	n.a.	0.60	0.60
<u>Scenario 3: Global trading is permitted</u>						
USA	-0.14	-0.14	-15.41	-21.91	-0.29	-0.33
Other Asia	0.25	0.36	1.33	2.63	0.06	0.02
China and India	0.34	0.52	4.55	13.58	-0.17	-0.07
Mexico and OPEC	-0.36	-0.42	1.40	3.02	-0.45	-0.48
Rest of World	0.03	0.08	2.42	5.02	0.06	0.04

Magnitudes of the measures are not directly comparable

As GDP and Direct Cost are annual measures they are presented in the above table for two different years (2010 and 2020), while EV and PVC are present value measures and hence a single number. The table makes it clear that the magnitudes of the welfare measures are not directly comparable. Aside from the differences in annual versus present value measures, EV, PVC and GDP are usually reported in terms of percent change relative to base case values, while direct cost can only be estimated in absolute monetary units (because there is no "base case" cost to which to compare them).¹⁷

Different policy rankings

In addition, the differing metrics suggest different policy rankings. For example, "Other Asia" would prefer global trading under the EV and PVC metrics, but not under the GDP metric (where the preference order is exactly reversed in 2010); further, this region would prefer Annex B trading if ranking were based on 2020 GDP.¹⁸ Similar ranking reversals also appear for "China and India". Finally, for "Rest of World", the preferred policy order based on EV is Global > Annex B > No ET; but based on PVC it is No ET > Global > Annex B; for 2010 GDP it is No ET > Annex B > Global; for 2020 GDP it is Annex B > No ET > Global.¹⁹ Few variants of the potential rankings are left.

The table above shows that in the first two scenarios EV, PVC and GDP provide inconsistent results for "Other Asia" and "Rest of World". Under EV, both "Other Asia" and "Rest of World" are shown to be worse off by policies that do not allow emissions trading, or that restrict emissions trading to Annex B countries. The GDP metric implies that these policies are beneficial for these regions. PVC also contradicts EV in most of these cases.

¹⁷ Alternatively, each of the other metrics could be presented in absolute terms, but those results were not originally reported.

¹⁸ This instability in the ordering based on GDP also highlights the problems of choosing which years to emphasize more when using a metric that is always presented on an annual basis rather than in present value.

¹⁹ Further, there is no discounting rule that could be applied to the GDP metric that would ever allow Global Trade to be the most preferred, even though Global Trade is the most preferred under the EV criterion. Thus, the altered rankings cannot be a mere by-product of attempting to compare a present value to annual values.

5 Conclusion

Our comparison of different welfare measures in the context of a single model that can calculate all shows that the different measures are not consistent. In fact differing metrics can be directly contradictory and produce different policy rankings.

Accepting the important value judgement that 'what the consumer prefers is better' a 'good' economic welfare indicator must be judged by whether or not it accurately reflects consumers' preferences. EV, which is defined as the amount of money that would have to be taken away from a consumer to put that consumer at the same level of utility as they would experience under the climate change policy, does this by construction. Therefore the EV ranking can be interpreted as showing that with appropriate compensation, the policy with the lowest loss in EV has the potential to make all members of society better off than any policy with a higher loss in EV. Other measures of impact, such as direct costs and GDP or GNP, cannot be guaranteed to always accurately reflect this same ordering of consumers' preferences.

Although EV is the most appropriate and ideal metric, there are issues associated with its estimation. Ideally, welfare should be measured over an infinite horizon - but models must be truncated at some finite point. If this is done at too early a point in time, or without a correction for post-horizon effects, the result will be a distorted measure of EV that does not properly rank policies. In particular, when two policies that lead to different levels of investment within some truncated time horizon are compared, it is necessary to include some approximation of the different values of capital left in place for future generations at the end of the modelling time horizon. The choices are to extend the time horizon until there is no significant change in results from further extension or to add an estimate of post-horizon welfare loss. Our experiments show that for a 30 year horizon, an estimate of post-horizon welfare loss is required. Unless this correction is made, calculations of welfare can be as misleading as any other measure.

Our next paper covers: (1) how to correct of after-horizon effects and (2) do other measures give ranking different from a correctly estimated EV?

Appendix

Mathematical derivation of equivalent variation and compensating variation

The compensation criterion tells us that a policy will increase the welfare of society as a whole if those who "win" from the policy change can compensate those who "lose". A "winner" is defined as a consumer who is able to achieve a higher level of satisfaction (or utility) under the new policy scenario. For this reason we would ideally like a welfare measure that can quantify the change in utility that arises as a result of the change in policy. The difficulty, however, is that utility is purely ordinal: that is, preferences are ranked, but the rankings do not indicate how great the differences in satisfaction are between alternative choices. One way of obtaining cardinal references to these rankings is to construct a series of budget constraints that calculates the minimum cost of reaching each level of satisfaction. Samuelson (1974) termed this the 'direct money metric utility function', where he defined the function to be equal to the minimum expenditure at prices p necessary to purchase a bundle at least as good as x (where x consists of a collection of goods). Formally this definition can be written as:

Equation A-1. Direct money metric utility function

$$m(p, x) \equiv e(p, u(x))$$

Where:

m	money metric utility function
p	price level
x	an allocation of goods
e	expenditure function
u(x)	utility (or consumer satisfaction level) associated with bundle x

The expenditure function $e(p, u(x))$ is strictly increasing in u and (if we fix p) thus becomes a monotone increasing function of utility (u) itself. For this reason the direct money metric utility function can be taken as a utility indicator in its own right.

Moving one step further: if we now allow prices to change to q , we can construct an 'indirect money metric utility function' that measures how much money one would need at prices p to be as well off as one would be under prices q and income m . An 'indirect utility function' (expressed as $v(p, m)$) is defined as the maximum utility achievable at given prices and income, and it can be shown that the 'indirect money metric utility function' is simply a monotonic transformation of that function.

The 'indirect money metric utility function' is given by the following equivalency and can be shown graphically, as in the diagram below:

Equation A-2. Indirect money metric utility function

$$u(p; q, m) \equiv e(p, v(q, m))$$

Where:	u	indirect money metric utility function
	p	initial price level
	q	new price level
	m	income
	e	expenditure function
	$v(q, m)$	indirect utility function

Figure A-1. Direct and indirect money metric utility functions (Varian, 1992)

The indirect money metric utility function gives us a cardinal ranking of preferences as prices and incomes change. Using this function we are able to directly measure the difference in utility (or satisfaction) that arises from the introduction of a new policy, as shown by the following.

Let us suppose that the current economic situation is characterised by prices p^0 and income m^0 , and that there is a climate change policy under consideration that would increase prices and income to p^1 and m^1 respectively. It can be shown that the change in utility that is attributable to this shift is simply the differences between the associated indirect money metric utility functions as demonstrated in the equation below:

Equation A-3. Utility difference between two policies using the indirect money metric utility function

$$\mu(q; p^1, m^1) - \mu(q; p^0, m^0)$$

There are in fact two answers to the above equation that depend on the value ascribed to q . q could be chosen so as to represent the base level prices, p^0 , or, alternatively, the future prices, p^1 . Both methods are equally appropriate and lead to the two well-known measures of welfare change, compensating variation and equivalent variation. The equations for each of these are laid out below:

Equation A-4. Compensating and equivalent variations

$$EV = \mu(p^0; p^1, m^1) - \mu(p^0; p^0, m^0) = \mu(p^0; p^1, m^1) - m^0$$

$$CV = \mu(p^1; p^1, m^1) - \mu(p^1; p^0, m^0) = m^1 - \mu(p^1; p^0, m^0)$$

As shown in the equations above the compensating variation uses the future prices associated with the policy scenario, p^1 , as a base. Thus the compensating variation can be defined as the income that compensates the individual for the proposed change from (p^0, m^0) to (p^1, m^1) at the new prices p^1 . By contrast, equivalent variation uses the current prices as a base and can thus be defined as the income that is equivalent to the proposed change from (p^0, m^0) to (p^1, m^1) at current prices p^0 . These welfare measures, and their relationship to prices p^0 and p^1 are shown in the diagram below:

Figure A-2. Equivalent and compensating variation

Compensating variation and equivalent variation each measure the change in utility that arises from a policy shift. Each is calculated directly from the money metric utility function, and requires only the quantities of goods consumed. They will not always be equal, however. Indeed the magnitudes of these measures may be quite different, as they are based on different price sets. Yet while the magnitudes may differ, the measures will never be in contradiction, so that the sign of each measure will always be an accurate representation of the utility difference.

The choice of which of these measures to use is somewhat arbitrary. However, equivalent variation is more appealing for comparing a range of policies for the following reasons:

- It measures the change in utility (and income) in current prices. This negates the need to estimate the likely price shift that results from a policy change.
- When comparing across a range of policies compensating variation requires the estimation of a set of new corresponding prices, whereas equivalent variation uses only current prices. The computational difficulties of predicting a range of possible prices makes the compensating variation more difficult to implement.
- Chipman and Moore (1980) show that compensating variation is an ordinal money metric only under the assumption that preferences are homothetic. EV has no such restrictions.