Pareto ordering and compensation

Mario Tirelli
April, 2010

1 Preliminary definitions

Economic fundamentals. Consider an economy with a finite number \( L + 1 \) of commodities and \( H \) individuals. A consumption profile for an individual \( h \) is a vector indicating the amount of each commodity to \( h \), \( x^h := (x^h_0, \ldots, x^h_L) \) in \( \mathbb{R}^{L+1}_+ \). An allocation \( x \) is a consumption profile for all the \( H \) individuals, in \( X = \Pi X^h \subset \mathbb{R}^{H(L+1)}_+ \). An allocation is feasible in an economy with aggregate resources \( X \) if it is attainable with the aggregate resources \( X \); i.e. if it belongs to the set \( A(X) = \{ x \in X : \Sigma x^h \leq X \} \).

The set of aggregate resources is denoted by \( X \).

Let \( \succeq^h \) be the weak preference ordering of individual \( h \), defined on \( X \). Assume \( \succeq^h \) is reflexive, transitive and complete. We can define the strict preference ordering \( \succ^h \), as follows: \( y \succ^h x \) iff \( y \succeq^h x \) and \( y \not\sim^h x \). Analogously, we can define the indifference relation \( \sim^h \) as: \( y \sim^h x \) iff \( y \succeq^h x \) and \( x \succeq^h y \). Denote \( \succeq := (\succeq^1, \ldots, \succeq^H) \) a preference profile. For any two allocations \( x, y \), read \( y \succeq x \) as \( y \succeq^h x \) for all \( h \).

For given individual preference profile \( \succeq \) we now focus on how these preferences can be aggregated into a “social preference”, denoted by \( \succeq_{so} \), to form a weak social ordering on \( \mathbb{R}^{H(L+1)}_+ \); where by social ordering we mean a preference ordering for the society as a whole. The social choice theory aims at studying preference aggregation rules (or social welfare function): mappings \( F \) from the domain of individual preference profiles to the domain of social preference orderings,

\[
(\succeq^1, \ldots, \succeq^H) \xrightarrow{F} \succeq_{so}
\]

\( F \) is required to possess some properties (or axioms). For example, Arrow (1951) requires that a social welfare function be transitive, complete, universal and observe the axioms of non-dictatorship and independence of irrelevant alternatives.\(^1\)

Social welfare function. Let us explain the meaning of these properties. Universal means that the social welfare function is well defined for any individual preference profile; that is, it delivers a social ordering for any community of individuals. Non-dictatorship means that there does not exist an individual \( h \) such that \( y \succ^h x \) implies \( x \succ_{so} y \), for any feasible pair \( x, y \), irrespectively of the individual preference orderings.

\(^1\)He did also require that a social welfare function is non to be imposed. A social welfare function is said to be imposed if, for some pair of alternatives \( x, y \), the social preference ordering remains the same irrespectively of the individual preference orderings. In other words, the social preference ordering is insensitive to the change in individual tastes and values.
of all the other individuals. Consider a social welfare function $F$, and two individual preference profiles $\succeq, \succeq'$, respectively, inducing two social orderings, $\succeq_S, \succeq'_S$; $F$ satisfies the **independence of irrelevant alternatives** when for any pair of alternatives $x, y$, $y \succeq_h x$ if $y \succeq'_h x$ for all $h$, implies the same social ordering between $x, y$ under $\succeq_S$ and $\succeq'_S$. In words, given a social ordering between any two alternatives $x, y$, a change of individual preferences (say a change in the community of decisors) which does not imply a change of the individual ordering between $x, y$, does not lead to a change in the social ordering of $x, y$.

There exist many social choice rules, aggregating individual preferences into social preference ordering, which are not to be considered social welfare functions, because they either fail to satisfy some desired axiom (e.g. Arrow’s or others) or produce a social ordering that fails to be transitive or complete. Indeed, Arrow’s impossibility theorem shows that with more than two alternatives ($X$ is made of more than two elements) a social welfare function (with Arrow’s properties) does not exist. Modifying the axioms of social choice leads to different conclusions. But these modifications are often questionable, or (in most cases) do not lead to revert Arrow’s theorem.

Finally, many largely used aggregation rules fail to satisfy Arrow’s axioms. The pairwise, complete, majority voting rule produces cycles, hence it induces a social ordering that is not transitive (Condorcet paradox). The unanimous voting rule induces a Pareto ordering that is transitive but fails to be complete. Using cardinal preference, like in the Borda count, restores completeness but violates the independence or irrelevant alternatives. Other directions to restore completeness, without imposing cardinality of preferences, are those based on the idea of potential Pareto principle and compensations.

In what follows we temporarily leave the Arrow’s axiomatic approach to social choice aside, and focus on the Pareto ordering. We do so, because the Pareto ordering captures the concept of efficiency, which is central in economics: every allocation that fails to be Pareto efficient entails some “wasteful” use of resources. The idea of interpreting inefficiencies as waste is immediately clear if we think at production-efficiency: an allocation is production efficient if it is not possible to rearrange the inputs across production plans and technologies such as to increase the output of some good without reducing those of the others.

## 2 Pareto ordering

We say that an allocation $x$ is **weakly Pareto superior** to an alternative $y$ if $x \succeq y$. Further, an allocation $x$ is **Pareto superior** to an alternative $y$ if $x \succeq y$ and $x \succ_h y$ for some individual $h$ (alternatively, $x \succeq y$ and $x \ngtr y$). Hence, an allocation $x$ in $A(X)$ is **Pareto optimal** with respect to $X$ if there does not exists an alternative $y$ in $A(X)$ which is Pareto superior to $x$.

The concept of Pareto superiority yields a social ordering, the Pareto ordering, which is reflexive and transitive, but fails to be complete. Namely, there are allocations which cannot be ordered according to this criterion. Examples can be derived both for pairs of allocations which lie on and off the set of Pareto optima.

**Problem 1** Take the Edgeworth box. Do you agree that allocations on the contract curve cannot be Pareto ordered? (Why?) Can you find two feasible allocations, one on the contract curve and the other not on the contract curve, which cannot be Pareto ordered?

One of the reasons why some outcomes cannot be Pareto ordered is that the Pareto criterion does not
take into account the "intensity" of individual preferences, i.e. by how much each individual $h$ like or dislike $y$ relative to $x$. If one had to bring preference "intensity" into the picture, it would be possible to think at situations in which by going form $x$ to $y$ those who have a welfare gain are so much happier than before, that in order to attain this transition would be willing to "compensate" the agents who have suffered a welfare loss. The Hicks-Kaldor criterion adds to the Pareto criterion the possibilities to implement such individuals' compensations. It is a remarkable criterion since it is formulated for ordinal preferences; its application does not require neither an inter-personal comparison of preferences nor cardinality at the individual level. We explain why in the next section 4, after having formally defined the criterion.

3 Potential Pareto ordering: the compensation principle

**Definition 2** (Kaldor-Hicks criterion) $y$ is Kaldor-Hicks (KH) socially preferred to $x$, $y \geq_{KH} x$, if there exists a $y' \in A(\Sigma y^h)$ such that $y' \succeq x$. $(y' - y)$ identifies a compensation profile.

This definition establishes a weak ordering, exploiting the notion of weak-Pareto ordering. An ordering can be similarly defined by using the notion of Pareto superiority: $x \succ_{KH} y$ if there exists a $y' \in A(\Sigma y^h)$ such that $y'$ is Pareto superior to $x$. This link between the compensation principle and the Pareto ordering explains why the first is often called potential Pareto ordering.

Since the ordering of any two alternatives $x, y$ depends on the individual preferences over a distinct alternative $y'$, the $KH$ ordering fails to satisfy the independence of irrelevant alternatives. To analyze the other properties of $KH$, first, suppose that there is just one possible aggregate bundle $X$, $\mathcal{X} = \{X\}$.

**Fixed aggregate resources $X$.** It is easy to show that $KH$ ordering is a weak ordering, which is reflexive, complete and transitive. For completeness, suppose two allocations are placed on the efficient frontier, then they can be obtain one from the other with the same aggregate resources without producing any strict Pareto improvement, hence they are $KH$ indifferent. Suppose they are as in figure 1; then they can clearly be ordered: any allocation on the frontier, like $y$, is $KH$ strictly preferred to any allocation in the interior, like $x$. Finally, consider two allocations $x, y$ which are in the interior of the set $A(X)$. Then, each allocation generates an attainable set, respectively, $A(\Sigma x^h)$ and $A(\Sigma y^h)$. Let the utility possibility set of $A(X)$ be the image of this set under $U = (u^1, ..., u^H)$, denoted $U(A(X))$. If preference are weakly

---

2Arrow proposed a slight modification of $KH$ compensation criterion such that the independence of irrelevant alternatives holds (see Arrow (1951), chapter IV, p.40). The idea is to divide the alternatives into mutually exclusive sets, each subset containing only those alternatives which can be obtained one from the other through compensations (i.e. redistributions of the same aggregate resources). Any pair of alternatives within the same set, in $KH$ spirit, may be regarded as made of equivalent allocations. That is, any of these, mutually exclusive, subsets forms a compensatory-equivalence class. Within each one of these subsets, the choice of one alternative over the other is just a matter of preferences over alternative distributions; hence, in Arrow’s words it is a matter of “ethical standards concerning distributions”. Thus, a social ordering could proceeds in two steps: first, order the allocations in each equivalence class according to ethical standards concerning distributions; then, having formed a new set of admissible alternatives, which do only belong to different equivalent classes (for different levels of aggregate resources), apply $KH$ ordering.

3Here we are assuming that preferences are non satiated.
increasing, $\Sigma x^h \leq \Sigma y^h$ implies $U(A(\Sigma x^h)) \subseteq U(A(\Sigma y^h))$, with equality if the two sets coincide. Therefore, $\Sigma x^h \leq \Sigma y^h$ implies $y \succeq_{KH} x$. Knowing this, it is also easy to show that $\succeq_{KH}$ is transitive [do it as an exercise].

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Figure 1}
\end{figure}

What if $\mathcal{X}$ is made of more than one element?

**Variable aggregate resources.** Suppose $\mathcal{X}$ is made of two elements, $\mathcal{X} = \{X, X'\}$. This is, for example the case in which there is a simple technology which allows to produce two levels of aggregate resources $X, X'$. To fix ideas, there are two goods ($L = 2$), leisure $l \leq 2T$ and a coconuts $c$, and the latter can be produced out of labor, in two levels so that, in the aggregate either 10 coconuts are collected using 1/4 of the total endowment of time of agents, or 20 coconuts using half of 2T: $X = (\frac{1}{4}2T, 10)$ or $X' = (\frac{1}{2}2T, 20)$. An allocation $x = (l, c)$, is a pair of leisure and coconut for each agent.

We can distinguish two cases:

1. One attainable set is included in the other: either $A(X) \subseteq A(X')$ or $A(X') \subseteq A(X)$.

2. None of the inclusions in 1. holds.

Case 1. can be dealt with just considering the previous discussion. You can reiterate your proofs of the properties of $KH$ simply considering $A(X')$ as the set of relevant alternatives.

Case 2. is more problematic, since then $KH$ yields a weak ordering that is neither transitive nor complete.\(^5\) Let us show this with an example. In figure 2 we have labelled some allocations. You should check

\(^4\) Just do it graphically. Fix the allocation of agent 1; hence her utility $u^1$. Give all the rest of aggregate resources to agent 2. Clearly, if the aggregate amount of commodities is greater under $\Sigma y^h$ than $\Sigma x^h$, the utility of agent $u^2$ will also be higher. This holds for every initial feasible allocation given to agent 1.

\(^5\) These criticisms were, first, raised by Scitovsky (1941).
that $KH$ yields the following cycle, hence that it is not transitive:

$$x_1 \succeq_{KH} x_2, x_2 \succeq_{KH} y_2, y_2 \succeq_{KH} x_1$$

This is a non-trivial cycle since the first two relations hold strictly (namely, $x_1 \succ_{KH} x_2, x_2 \succ_{KH} y_2$). Finally, $KH$ is not complete, since it fails to order $x_2$ and $y_2 : x_2 \succ_{KH} y_2$, and $y_2 \succ_{KH} x_2$, since after a redistribution—one finds feasible allocations which yield (strict) Pareto improvements.

Observe that what is peculiar about $KH$ ordering between two allocations $y, x$ is that one search for a potential Pareto improvement on a given allocation $x$, not just over an alternative $y$, but over the whole set of allocations which are attainable through $\Sigma y^b, A(\Sigma y^b)$. We could restate our definition as follows: $A(X')$ is Kaldor-Hicks socially preferred to $x$, $A(X') \succeq_{KH} x$, if there exists a $y' \in A(X')$ such that $y' \succeq x$.

The fact that allowing for compensations induces to order sets, rather than just allocations, is even more striking in the compensation principle proposed by Samuelson (1950).

**Definition 3** (Samuelson social ordering) $A(X)$ is Samuelson ($S$) socially preferred to $A(X')$, $A(X) \succeq_S A(X')$, if for all $x \in A(X')$ there exists a $y \in A(X)$ such that $y \succeq x$.

Samuelson social ordering says that $y \in A(X)$ is socially preferred to $x \in A(X')$ if and only if for every $\tilde{x} \in A(X')$ there exists a $\tilde{y} \in A(X)$ that is Pareto superior to $\tilde{x}$. This is equivalent to say that the Pareto frontier induced by $A(X)$ contains the one induced by $A(X')$. The latest is as in case 1. above. Therefore, as for $KH$, in this case the ordering is transitive and complete. By contrast, for sets as those in case 2. (see figure 2), Samuelson’s is of no help; indeed, it is unable to order sets, hence it is incomplete.

It is immediate to relate Samuelson’s with Kaldor-Hicks: $y \in A(X)$ is $S$—socially preferred to $x \in A(X')$ if and only if it is $KS$ preferred to every allocation $\tilde{x} \in A(X')$.

In the remaining sections of this document we show how Kaldor-Hicks principle is typically implemented. Extensions to Samuelson’s are straightforward and therefore omitted.
4 Implementation and testing of Kaldor-Hicks’

Most economists identify the Kaldor-Hicks principle with a specific criterion: the aggregate compensating variation (or equivalent variation), widely used in cost-benefit analysis. In this section we define and characterize this criterion using the concept of individual willingness-to-pay (w.t.p.). Our formulation is conceptually similar to the compensating variation, except that compensations are defined in allocation space, rather than in income space. More precisely, we define the w.t.p. for an allocation $y$ at the status quo allocation $x$ as the maximum amount of a commodity, say the first commodity, the individual is willing to forgo (or requires as compensation) to achieve $y$. We label and interpret the first commodity as the numéraire, and assume it is common to all agents.\textsuperscript{6}

4.1 Definition of willingness-to-pay

In our definitions, compensations are left unrestricted; namely they can be implemented as transfers between consumers in any commodity. For simplicity, we now restrict compensations to be carried out in one commodity only, the numéraire commodity, the first commodity labelled $l = 0$, and that every individual $h$ cares only about his/her own consumption. $x^h := (x^h_0, ..., x^h_{L+1}) \in X^h = \mathbb{R}^{L+1}_+$; to partition goods between numéraire and non-numéraire, we use the notation $x^h := (x^h_0, x^h_1) \in \mathbb{R}_+ \times \mathbb{R}^L_+$.

Suppressing indexes, the individual willingness-to-pay (w.t.p.) for $y$, in terms of the numéraire good, at $x$, is by definition the maximum numéraire quantity $w \in \mathbb{R}$ the individual is willing to give up, or requires, in order not to suffer a welfare loss:

$$\max w \ s.t. \ y - (w, 0) \succeq x$$  

where $0 := (0, ..., 0)$ is an $L$-vector of zeros. Fixing the status quo, $x$, this defines a function of $y$ defined at $x$, $w = w_x(y)$ whose argument we occasionally omit. Observe that $w$ may be negative, interpretable as a compensation to be received. Indeed, if $h$ is worse off with $y$ than with $x$, $w_x(y)$ is the (least) compensation $h$ requires to give up $x$ for $y$. Since payments are out of the promised numéraire, then $w_x(y) \leq y_0$.

To check if w.t.p. is well defined, we first have to verify that the set of $w$ such that the constrained $y - (w, 0) \succeq x$ is nonempty. Preference completeness implies that we have to possible cases, either $y \succeq x$ or $x \succ y$. The first case is trivial, since $w = 0$ is a feasible choice. If $x \succ y$ we have to restrict preferences to be such the individual has a ”sufficiently strong” taste for the numéraire: i.e. there exists a compensation $w$ such that there exists $y - (w, 0) \succ x$. This holds for preferences which do not have an asyntotus (along the 0-axis) with respect to the numéraire commodity. A technical assumption of 0-desiderability that works is: for every $x \succeq 0$ and $y = (y_0, y_1) \succ 0$, there exists a sequence $y^n := (y^n_0, y) \succ 0$ and a $\bar{n}$ such that $y^n \succeq x$ for all $n \geq \bar{n}$.\textsuperscript{7} Therefore, the w.t.p. is well defined as a solution to problem (1).

Further, suppose that preferences are Inada ($-w + y_0 > 0$) and 0-continuous ($y \succ x$ implies that there exists a scalar $\varepsilon > 0$ such that $y - (\varepsilon, 0) \succ x$). Then, a solution $w$ to problem (1) satisfies $y - (w, 0) \sim x$.\textsuperscript{8}

---

\textsuperscript{6}The choice of the numéraire is clearly arbitrary. Another alternative would be to assume that individual compensations are quantified as a fraction $0 \leq \alpha \leq 1$ of every commodity, $\alpha y$ being individual-specific.

\textsuperscript{7}Observe that, for any $x, z$, $(x_0, x_1 + z) \succ 0$ the relevant sequence is $y^n := (x_0 - w^n, x_1 + z) \succ 0$.

\textsuperscript{8}To check this statement, suppose (by contradiction) that $y - (w, 0) \succ x$, then by continuity there exists a scalar $\varepsilon > 0$, $y - (w + \varepsilon, 0) \succ x$; hence, $w := w + \varepsilon > w$ is feasible payment and is greater than the maximum $w$; a contradiction.
Finally, if preferences are 0-increasing \( (x - (\varepsilon, 0) \succ x, \text{ for all } \varepsilon > 0) \), then the solution to problem (1) is unique.\(^9\)

To summarize,

**Proposition 4** Suppose preferences satisfy 0-desiderability, Inada, 0-continuity and 0-increasing. Then for every pair of allocations \((x, y)\) the willingness-to-pay \(w_x(y)\) is defined as the solution of problem (1) and the constraint

\[
y - (w, 0) \succeq x
\]

holds with indifference at \(w_x(y)\). Furthermore \(w_x(y)\) is a unique maximum.

Thus, under reasonably mild assumptions we can define the individual willingness-to-pay for \(y\) at \(x\) as the numéraire quantity \(w\) that satisfies

\[
y - (w, 0) \sim x
\]

where \(w\) is a payment when the change from the status quo is welfare improving for the individual and a compensation when it is welfare diminishing. Graphically, in a two-commodities economy, the willingness-to-pay is equal to the line segment on the horizontal axis marked with \(w\) in the following figure,

\[\text{Notice that, unlike in our figure, our characterization of w.t.p. neither requires convexity of preferences, nor assumed non-satiation in any of the commodities other than the numéraire; thereby including cases in which there are ‘bads’, and externalities. However, a Inada is introduced to rule out situations in which the numéraire good to be consumed by an agent, net of the payment, be negative (something that would happen if in the figure the indifference curve through } x \text{ crosses the vertical axis below } y_1.\]

\(^9\)Suppose not, there are two solutions \(w, w'\). Then, by completeness of the reals, either \(w \geq w'\) or \(w < w'\). Strict inequalities would contradict the fact that they are both maxima.
Remark 5 (willingness to pay versus compensating variation and equivalence variation) The fact that we choose good 1 as the numéraire is arbitrary. We could propose alternative formulations such as a uniform destruction of all goods, or such as one that reasons in terms of indirect utilities and income compensation. Yet, for the latter we should adopt some system of good princes.

The concept of willingness to pay is related to those of compensating variation (CV) and equivalence variation (EV). These can be generally expressed as follows. When an agent experiences a change from a situation–status quo to another–new status–, the CV is the payment the individual must make or accept to maintain the same status quo utility; while the EV is the payment the individual must make or accept to achieve the utility corresponding to the new status. These concepts are usually applied to evaluate the effects of changes in the parameters of a consumer problem (prices, income, tax/subsidies, etc.). In the canonical textbook illustration (e.g. see Varian's Microeconomic Analysis), preferences are represented by a money metric indirect utility function, \( \mu(p'; p, m) \) defined as the minimum income needed at \( p' \) to be able to reach an (indirect) utility \( v(p, m) \). Here \( (p, m) \) is the status quo and \( (p', m) \) is the new status \((m\) does not vary). If the change in prices \( p \to p' \) is beneficial for the agent, i.e. \( v(p', m) - v(p, m) > 0 \), then the compensating variation is the value \( CV := \mu(p; p, m) - \mu(p'; p, m) = m - \mu(p'; p, m) > 0 \); otherwise, if \( v(p', m) - v(p, m) \leq 0 \), \( CV \leq 0 \). Conversely, the equivalence variation, \( EV := \mu(p; p', m) - \mu(p'; p', m) = \mu(p; p', m) - m \); which is again positive or negative depending on whether or not the change is beneficial for the individual. For quasilinear preferences \( (v(p, m) = m + k(p)) \) the two measures are identical (also equal to the consumer surplus). This is intuitive since now, by definition, \( v(p, m + EV) - v(p', m) = 0 = v(p', m - CV) - v(m, m) \), which holds iff \( k(p') - k(p) - EV = k(p') - k(p) - CV \).

4.2 Properties of willingness-to-pay

We now characterize the willingness-to-pay as a utility index.

Lemma 6 \( w_x(y) \geq 0 \) if and only if \( y \gtrless x \).

Proof. (necessity) By contradiction, suppose \( x \gtrless y \), and show that this implies \( w_x(y) < 0 \). Since, by proposition 4, \( w \) satisfies (2) with indifference, \( w \) is such that \( y - (w, 0) \sim x \). Since \( x \gtrless y \) the latest implies \( y - (w, 0) \gtrsim y \), which–because preferences are increasing in the numéraire good–can only hold if \( w < 0 \), yielding a contradiction. \( w_x(y) \geq 0 \Rightarrow y \gtrsim x \).

(sufficiency) Suppose \( y \gtrsim x \). Again, by definition, \( w \) satisfies \( y - (w, 0) \sim x \). But since preferences are increasing in the numéraire good, \( w = w_x(y) \geq 0 \). ■

We can use lemma 6 to establish the following.

Proposition 7 The individual willingness-to-pay represents a weak preference ordering, reflexive, transitive, complete.

4.3 The ‘popular’ Kaldor-Hicks criterion

Let us assume that \( \mathcal{X} = \{\omega\} \), \( \omega \in \mathbb{R}^{L+1}_{++} \) denotes the total resources available in the economy, also partitioned as \( (\omega_0, \omega_1) \). An allocation \( x \) is feasible if it is an element of \( A(\omega) \).\(^{10}\) Since the individual willingness-to-pay

\(^{10}\)Here, we would like to keep on considering the possibility that at a Pareto efficient allocation, not all the resources are distributed; either because some commodity is a ‘bad’, or because it induces satiation.
are all expressed in units of the same good, by hypothesis, the numéraire good, we can also define a measure of aggregate willingness-to-pay for a society: the aggregate willingness to pay of the society for $y$, at $x$, is $W_x(y) := \sum_h w_h^b(y)$. Being the sum of individuals’ willingness-to-pay, $W_x(y)$ is defined for ordinal preferences. Moreover, under the above assumptions, it is well defined and bounded above by $\omega_0$.\(^{11}\)

The following proposition establishes that we can use this measure to test for potential Pareto improvement.

**Proposition 8** For any two alternatives $(x, y)$, $W_x(y) \geq 0$ implies that $y \succeq_{KH} x$; that is, there exists a $y'$ that can be derived from $y$, through compensations in the numéraire, such that $y'$ is Pareto superior to $x$.

**Proof.** Let us show the implication for a strict inequality, $W_x(y) > 0$, implying there is at least an individual, say $h$, with $w_h^b(y) > 0$. Let $y' := y - (w_x(y), 0)$. By definition of w.t.p., $y' \sim x$. Consider $h$; by local non satiation and 0-increasing, for every $\epsilon > 0$ there is a $y^h + (\epsilon, 0) \succ y^h$. Choose $0 < \epsilon \leq W_x(y)$. Then, by transitivity $y^h + (\epsilon, 0) \succ x^h$. Therefore, $y^h + (\epsilon, 0)$ and $y^i$ for all $i \neq h$ is feasible: $y'_i = y_i$ for all $l \neq 0$, and

$$\Sigma y^l' + \epsilon = \Sigma (y^l_0 - w^l(y)) + \epsilon = \Sigma y^l_0 - W_x(y) + \epsilon \leq \Sigma y^l_0 - W_x(y) + W_x(y) \leq \omega_0.$$ 

Moreover, it is Pareto superior to $x$, hence $KH$–superior to $x$. \(\blacksquare\)

We can do more than this by establishing a partial converse of this proposition.

**Proposition 9** For any two feasible alternatives $x, y$, $y \succeq_{KH} x$ implies there exists a $y'$ that can be obtained from $y$, through compensations, such that $W_x(y') \geq 0$.

**Proof.** By definition, if $y$ is $KH$ superior to $x$, then there exists a $y'$ that can be derived from $y$ through compensations such that $y' \succeq x$. By definition of w.t.p., $y^h \succeq_{KH} x^h$ implies $w_h^b(y') \geq 0$. Aggregating, $\Sigma_h w_h^b(y') =: W_x(y') \geq 0$. \(\blacksquare\)

**Remark 10** (Ordinality of the measure $W$) To argue that the Kaldor-Hicks criterion $W$ is ordinal it suffices to establish the ordinality of the individual w.t.p.. To this end, if lemma 6 was not convincing, let us add the following observations. Assume that some individual preferences are represented by a utility function $u(\cdot)$, and take any arbitrary pair of allocations $x, y$. By proposition 4, $w_x(y)$ is such that $u(y - (w_x(y), 0)) = u(x)$. The w.t.p. is ordinal if $w_x(y)$ continues to make the latest hold with indifference for any arbitrary monotonic transformation $f : \mathbb{R} \to \mathbb{R}$ of the utility function. Let $w^*$ be such that $f(u(y - (w^*, 0))) = f(u(x))$. Then, since $f$ is injective, and the w.t.p. is unique, $f(u(y - (w^*, 0))) = f(u(y - (w_x(y), 0))) = f(u(x))$ if and only if $w^* = w_x(y)$.

**Remark 11** (No inter-personal comparisons) The Kaldor-Hicks criterion $W$ is constructed without making any inter-personal comparison of individual preferences, precisely, in the spirit of its proponents. Indeed, compensations, in form of individual w.t.p. are defined individual-by-individual and then aggregated to determine $W$. Therefore, even if agents have different ‘utility indices’, or if their indices are monotonically transformed, using different transformations, this does neither affect the individual w.t.p., nor $W$.

\(^{11}\) $y^l_0 - w^l(y) \geq 0 \Rightarrow \sum_h w_h^b(y) \leq \sum_h y^l_0 \leq \omega_0$
4.4 Criticisms

The criticisms we shall bring up are raised to the Kaldor-Hicks principle implemented using willingness-to-pay, also under different metrics typically used, compensating and equivalence variation, uniform collection of goods. Remember that all these metrics coincide if we compare allocations which are arbitrarily closed one another, i.e. they coincide at the margin. We recall some of the main criticisms next.

4.4.1 Arbitrariness

The fact that we choose the first good as the numéraire is arbitrary. We could propose alternative formulations, also implementing the Kaldor-Hicks principle, just based on the selection of a different numéraire, or on a uniform payment over all goods, or based on income. If income is taken as a measure of w.t.p., one should in general make clear which evaluation criterion or price system to adopt. For any discrete change of the status quo $x$ into the alternative social state $y$, it is natural to think that a competitive price system at $x$ will be different from the one prevailing at $y$. This is indeed the distinctive feature of compensating variation and equivalence variation.

4.4.2 Evaluating and implementing compensations

One of the main criticisms of Kaldor-Hicks is the actual difficulty of implementing compensations. Compensations are different, by sign and absolute level, across heterogenous individuals, and can only be computed having knowledge of individual preferences. Suppose, compensations should be implemented by a policy maker who knows only the distribution of individual characteristics, preferences and initial allocations, $(\preceq, x)$, but cannot identify individual types. Then, this policy maker can certainly compute all the willingness-to-pay, but faces problems in collecting any positive payments $w$ out of a proposed allocation $y$ that is not strictly Pareto superior to the status quo $x$.

A fundamental question is then: is it really necessary to implement compensations? Kaldor's answer is not really, since a state $y$ that is Kaldor-Hicks preferred to $x$ is one with a higher aggregate income, regardless compensations are actually carried out. Thus, if one abstracts from distributional issues, it is possible to conclude that a state with an higher aggregate income should be socially preferred to one with a lower aggregate income. However, unless one considers the trivial case in which $y$ is Pareto superior to $x$, implementing compensations should be taken seriously. Let us clarify these points establishing a few interesting properties of Kaldor-Hicks criterion.

**Fact 12** Consider two alternatives $y, x \succ 0$, and interpret $x$ as the status quo consumption allocation attained at a Walrasian equilibrium with prices $p$.

**Fact 13** $W_x(y) > 0$ only if $p\Sigma y^h > p\Sigma x^h$.

---

12 The uniform collection of goods is substantially different from Debreu's (1951) coefficient of resource utilization. It would be based on a coefficient computed for every agent, rather than a uniform coefficient computed with respect to aggregate resources. Indeed, Debreu's asks what is the maximum amount of resources that can be given up by the society to achieve a Pareto efficient state, $\Sigma y^h$, from a state $\Sigma x^h$, so that consumers are weakly better off with $y$ than $x$. 

10
Proof. \( W_x(y) > 0 \), implies that for all \( h \) there exists a compensation \( w^h = w^h(y) \), such that \( y^h + (-w^h, 0) \succeq_h x^h \). If individuals’ preferences are non-satiated in at least one commodity, the latest implies that \( p(y^h + (-w^h, 0)) \geq px^h \). Aggregating, \( \Sigma y^h + W_x(y) \geq \Sigma px^h \), yielding \( p\Sigma y^h > p\Sigma x^h \). ■

Notice that the reverse of this fact does not hold in general, unless marginal changes of \( x \) are considered (\( y \approx x \), i.e. \( y \) in a small neighborhood of \( x \)). Indeed, in the latter case, given ordinal utilities \( u \) representing preferences \( \succeq \), \( u(y) - u(x) = (1, \nabla_x) \cdot (y - x) \), where \( \nabla_x = (\partial_1 u/\partial_0 u, \ldots, \partial_L u/\partial_0 u) \) are marginal rates of substitutions.\(^{13}\) In a Walrasian equilibrium, normalized prices satisfy \( p = (1, \nabla^h) \) for all \( h \). Hence, \( 0 < p\Sigma y^h - p\Sigma x^h = \Sigma(1, \nabla^h)(y^h - x^h) \). Next, I can choose a (marginal) redistribution of \( y, y' := (y^h - (w^h, 0))_h \), such that \( (1, \nabla^h)(y^h - (w^h, 0) - x^h) = 0 \) for all \( h \). \( w^h = w^h_1(y) \) since \( 0 = (1, \nabla^h)(y^h - (w^h, 0) - x^h) = u^h(y^h - (w^h, 0)) - u^h(x^h) \). Moreover, it yields a positive aggregate willingness to pay: \( 0 = \Sigma(1, \nabla^h)(y^h - (w^h, 0) - x^h) = \Sigma(1, \nabla^h)(y^h - x^h) - \Sigma w^h \), implying \( \Sigma w^h = \Sigma(1, \nabla^h)(y^h - x^h) = p\Sigma y^h - p\Sigma x^h > 0 \).

Fact 14 \( y \) is Pareto superior to \( x \) then \( W_x(y) > 0 \).

Proof. Suppose \( y \) is Pareto superior to \( x \). Then, for each and every \( h \), \( y^h \succeq_h x^h \), strict for some individual. The latter implies that \( w^h_1(y) \geq 0 \) for all \( h \), strict for those individuals who strictly prefer \( y \) to \( x \). ■

The reverse of this fact is not true, in general. In other words, if one observes either \( W_x(y) > 0 \) or \( p\Sigma y^h > p\Sigma x^h \), it is only possible to conclude that \( y \succeq_{KH} x \), that is there exists compensations such that \( y \), corrected by these compensations, yields a new allocation \( y' \) that is Pareto superior to \( x \). This result clarifies why the problem of implementing compensations cannot be eluded.

A related, important, point in the previous discussion regards the choice of prices, \( p \). As Kaldor and Hicks themselves recognize their ordering of social states leaves distributional issues aside. This is sort of problematic in economies with multiple goods, since different distributions of resources imply different price systems. The only exception is when individuals have constant marginal utilities of income–as assumed by Kaldor and Hicks. Therefore, in general, it is not clear how one can effectively use the compensation principle to address efficiency completely disregarding distributional issues.

Finally, our previous discussion relies on the assumption that the economy has complete markets and thus individuals marginal rates of substitutions are equalized to relative prices at equilibrium. On the contrary this property does not, typically, hold if markets are incomplete, whether because of externalities, asymmetric information, transaction costs, trade limits. In deriving fact 13 one should account for the fact that if \( p \) is an equilibrium price of an incomplete markets economy, \( y^h \succeq_h x^h \) implies \( \nabla^h y^h \geq \nabla^h x^h \), with the vector of marginal rate of substitution of each commodity with the numéraire, \( \nabla^h \), being typically different from \( p \).

4.4.3 Scitovsky’s paradox, lack of completeness and transitivity

Scitovsky (1941) pointed out that the Kaldor-Hicks criterion violates two fundamental axioms of any weak social ordering: completeness and transitivity. This problem has been illustrated above, using figure 2.
and is a direct consequence of aggregation; indeed, individual preferences are assumed to be transitive and complete. Since, individual preferences are represented by the utility index based on the w.t.p., the latest inherits the properties of individual preferences. By contrast, when we aggregate w.t.p., $W$ may fail to be transitive and complete. For example, it may happen that

$$W_x(y) > 0 \text{ and } W_y(x) > 0$$

So if $W_x(y) > 0$ $y$ is socially strictly preferred to $x$, hence if the ordering were complete it would not be possible for $x$ to be socially preferred to $y$, which instead is implied by $W_y(x) > 0$. This also implies that transitivity would be violated.

Observe that this paradox vanishes if the measure of individual compensations is symmetric, namely when the compensation from $x$ to $y$ equals the compensation from $y$ to $x$, with the sign reversed. As illustrated in the figure, this is not the case, in general, for the w.t.p.\(^\text{14}\) The willingness-to-pay is not a Scitovsky-compensation. To remark how ‘disturbing’ this is observe that it implies that by repeatedly going back and forth from $x$ to $y$, a policy maker could extract from a single individual—and possibly from the society as a whole—an increasing amount of income or numéraire commodity, thereby activating a ‘money pump’:

$$w_x(y) + w_y(x) > 0 \quad (4)$$

Parametrically, consider an individual with preferences represented by $u(x_0,x_1) = x_0x_1$ and let $x = (1,2), y = (8,1), u(y) = 4u(x) = 8, w_x(y) = 6, w_y(x) = -3$.

\(^{14}\)An important exception is when preferences are quasi-linear in the numéraire, $u(x) = \alpha x_0 + u(x_1)$.
4.4.4 Boadway Paradox

Boadway first noticed that the aggregate willingness-to-pay to move from a status quo $x$ that is Pareto efficient, to another Pareto efficient allocation $y$ may be always non-positive and is strictly negative, under convex preferences (e.g. see the following figure).\textsuperscript{15} This paradox implies that having a negative Kaldor-Hicks index $W_x(y) < 0$ is only a necessary condition for $x$ to be Pareto superior to $y$; indeed, as in the picture, one can have $W_y(x) < 0$.

**Example 15** Parametrically, consider a Edgeworth box $3 \times 3$, and assume that individuals have identical preferences represented by $u(x_0, x_1) = x_0 x_1$. Consider two Pareto efficient allocations $x = ((2, 2), (1, 1))$ and $y = 31 - x$. $W_x(y) = W_y(x) = -1.5$.

Again, this is not in contrast with our characterization in proposition 7: if both $x$ and $y$ are Pareto optimal, our lemma does not say anything about the sign of the aggregate willingness-to-pay. This is the consequence of aggregation and to the fact that the w.t.p. is an asymmetric measure—in the sense that (4) may not hold with equality.

5 References


\textsuperscript{15}As for compensating and equivalent variations, an equivalent result can be shown by comparing two different Walrasian equilibria and exploiting the two welfare theorems. The Boadway Paradox is known since 1974, see Boadway (1982); see also the review article by Blackorby and Donaldson (1990).


