On Measuring Welfare ‘Behind a Veil of Ignorance’

Moti Michaeli

Abstract

In this paper we present two complementary methods for ranking distributions of incomes using the preferences of a large set of utilities that represent the members of society needing to reach a decision ‘behind a veil of ignorance.’ One method requires unanimous acceptance of candidate income distributions by all members of society, while the other requires anything but unanimous rejection. Two ordinal and complete rankings (corresponding to two distinct measures of welfare) emerge as a result of applying these methods: unanimous acceptance leads to the Rawlsian principle of maximin, while unanimous rejection leads to ranking income distributions according to their geometric means.

1 Introduction

Rawls (1971) raises the fundamental question of how we should choose to distribute incomes in society, or how we should rank such income distributions, “behind a veil of ignorance.” The idea of being ‘behind a veil of ignorance’ is that of a two-stage process, where in stage 2 an income would be *randomly assigned* to the decision maker according to the distribution of incomes he picked in stage 1. In economic terms, the question of ranking income distributions 'behind a veil of ignorance' is therefore similar to the question of ranking lotteries, i.e., distributions over final outcomes. Rawls then promotes ranking according to the *maximin criterion*, i.e., distribution A would be ranked higher than distribution B if and only if $\min_{a_i \in A} a_i > \min_{b_i \in B} b_i$.[1] An economist would probably approach the question quite differently: given utility function u, distribution A should be ranked higher than distribution B if and only if $E_A[u(a_i)] > E_B[u(b_i)]$. However, how can one choose the "right" utility function according to which to rank the distributions? In this paper we present two complementary methods for ranking such income distributions using the tendency of a large set of utilities to either accept or reject the lotteries corresponding to these distributions. This set of utilities is thought of as representing the "members of society" needing to reach a decision 'behind a veil of ignorance.' Two ordinal rankings emerge as a result of applying these two methods. One ranking is based on the Rawlsian principle of maximin, while the other ranks income distributions according to their geometric means. The essence of the idea is to borrow concepts that are used for ranking risky prospects and apply them in order to rank income distributions.

The question of how to reconcile considerations of income inequality with considerations of aggregate or average income goes back at least to Sheshinski (1972). Sheshinski

[1] We assume throughout that income distributions are bounded and have finite support in R^+, and therefore the maximum and minimum incomes exist and are well defined. We denote the incomes in the support of distribution A by a_i, $i = 1, 2..$ (and similarly for B).

develops a set of welfare functions from the Gini Index, and many would follow his footsteps and develop sets of welfare functions out of measures of inequality aversion (e.g., Ebert 1987). The drawback of this approach is that it often leads to a whole family of welfare functions and eventually to partial orders, while a unique and complete ranking (corresponding to a unique ordinal measure of welfare) is usually desirable. Shorrocks (1983) was able to partially reduce the incompleteness of rankings by using *generalized Lorenz curves*, that incorporate both inequality aversion and expectation, and to supply non ambiguous ranking for 84% of the pairs of economies he tested.

The tension between average income and income inequality when considering income distributions resembles the tension between expectation and risk when considering lotteries. It is therefore not surprising that the connection between inequality and risk (or sometimes welfare and risk) has already been indicated in the past, e.g., in Harsanyi (1955).[2] As Carlsson et al. (2005) note, the implication for the choice 'behind a veil of ignorance' was also considered in this context: "(l)oosely speaking, the more concave the utility function, the larger the relative risk aversion, implying that an individual choosing between different societies behind a veil of ignorance would be willing to trade off more in terms of expected income in order to achieve a more equal income distribution." Indeed, the most popular measure of inequality aversion, the Gini Index, resembles the most popular measure of risk, the variance, in that both are in fact measures of *dispersion*, ignoring the expected income in the former case and the expected return in the

[2] An interesting result in this context is that of Safra and Segal (1998), who combine constant risk aversion with other axioms to obtain a functional form that is a weighted average of the expected value functional, and the Gini inequality index.

latter. Accordingly, a measure of *welfare* that ranks income distributions not only by their dispersion but also by their expectations, can be compared to a measure of *riskiness* that ranks risky prospects or lotteries not only by their dispersion but also by their expectations. Such a measure of riskiness is that of Foster and Hart (2009): it preserves stochastic dominance (whereas variance is guaranteed to do so only for prospects having the same expected value), and it invokes a ranking of risky prospects that has a straight forward interpretation in terms of preferences and utilities (see Hart 2011). We therefore adopt (with some modifications) this new approach of Hart (2011) and use it in order to develop *complete* rankings of income distributions (corresponding to measures of welfare) that take into account both average income and income inequality in a manner that is determined by the utilities in the set of utilities that represents "society."

2 Utilities

Who are the members of society that will rank income distributions 'behind a veil of ignorance'? We impose four fundamental restrictions over the set of utilities U^* that represents those members:

(1) "More is (always) better": $\forall u \in U^*$, $u' > 0$.

(2) Risk aversion: $\forall u \in U^*$, $u'' < 0$. When considering income distributions, this implies some basic form of inequality aversion: every member of U^* prefers distribution A over distribution B if A second-order stochastically dominates B.

(3) Weakly increasing relative risk aversion: $\forall u \in U^*$, $\forall w > 0$, $\gamma_u(w) \equiv -\frac{wu''(w)}{u'(w)}$ is weakly increasing in w. This property, perpetuated by Arrow (1965, 1971), says that

acceptance of risky prospects is weakly decreasing with *relative* wealth. Informally, this means that the inclination to invest one's money in a risky prospect whose returns are given in *proportional terms* (like a stock) weakly declines with the amount of one's money.[3] The set of utilities that respect this property include all CRRA utilities (for which relative risk aversion is constant) and all CARA utilities (for which relative risk aversion is linearly increasing with wealth).[4]

(4) "strong aversion to bankruptcy": $\forall u \in U^*$, $\lim_{w \to 0^+} u(w) = -\infty$. This property says that being left with absolutely nothing is considered unacceptable by any member of U^*. It is a much weaker version of the concern raised by Rawls about the possibility of being the poorest member of society (and it is also expressed in absolute terms and not in relative terms, as in Rawls 1971).

3 Ranking methods

Ideally, we may have preferred to let the members of society, i.e., the set of utilities U^*, to rank each pair of income distributions by voting for their preferences, and then aggregate the votes in some reasonable way (e.g., by a majority rule). However, this procedure would lead to a conflict of interest for most ranked pairs (in fact, unanimity is guaranteed only when there is second order stochastic dominance), and it is impossible

[3] So, e.g., a person would be willing to invest all his money in a stock that yields +30% or -10% with equal probabilities if his initial wealth equals \$1000, while refusing to invest all his money in the same stock if his initial wealth equals \$1,000,000.

[4] In fact this set includes (but is not restricted to) all HARA utilities (for which CRRA and CARA are special cases), i.e., utilities that can be represented by $u(w) = \frac{\gamma}{1-\gamma}\left(\frac{w}{\gamma}+b\right)^{1-\gamma}$ for some parameters γ and b, s.t. $\frac{w}{\gamma}+b>0$. It is easy to verify that the relative risk aversion of a HARA utility is $w\left(\frac{w}{\gamma}+b\right)^{-1}$, and it is increasing in w $\forall w>0$.

to define a majority rule for this infinite set of utilities (notice that risk aversion is not bounded from above). To overcome this problem, let us replace voting between two income distributions with voting for or against any given income distribution (in a manner that will become clear shortly), and let us confine ourselves to veto-based decision rules. The right of veto can be implemented in one of two ways: giving the members of society a right to veto *acceptance* of an income distribution implies that the income distribution is said to be "accepted" by the members of society if and only if *every member accepts it*; giving them a right to veto *rejection* of an income distribution implies that the income distribution is said to be "rejected" by the members of society if and only if *every member rejects it.* These two ways of implementation are the two methods leading to the two distinctive measures of welfare we present here. But what does it mean that a member of society "accepts" or "rejects" an economy? Obviously, an outside alternative must be introduced. Let us assume for a moment that the outside alternative for every member in U^* is set to be some guaranteed income w_0. Then every individual with $u \in U^*$ is said to *accept* income distribution A at wealth w_0 if and only if $E_A[u(a_i)] > u(w_0)$.[5] Consequentially, we can say that income distribution A is *unanimously accepted* at wealth w_0 if A is accepted by *all* utility functions $u \in U^*$ at w_0, and that income distribution A is *unanimously rejected* at wealth w_0 if A is rejected by *all* utility functions $u \in U^*$ at w_0.[6] Thus, unanimous acceptance is required in order to say that an income distribution is accepted when members are given the right to

[5] One may replace the strong inequality with a weak one without changing the results of the model.
[6] The concept of *unanimous rejection* is analogous to the concept of *uniform rejection* in Hart (2011).

veto acceptance, whereas unanimous rejection is required in order to say that an income distribution is rejected when members are given the right to veto rejection. To dispose of the dependency on the specific wealth w_0, we use the following definitions.

We say that income distribution A *unanimously-acceptance dominates* income distribution B, which we write $A \geq_{UA} B$, if the following holds:

For every wealth level $w > 0$,

if B is accepted by all $u \in U^*$ at w

then A is accepted by all $u \in U^*$ at w.

Similarly, we say that income distribution A *unanimously-rejection dominates* income distribution B, which we write $A \geq_{UR} B$, if the following holds:

For every wealth level $w > 0$,

if A is rejected by all $u \in U^*$ at w

then B is rejected by all $u \in U^*$ at w.

In words, income distribution A unanimously-acceptance dominates income distribution B, if members are given the right to veto acceptance (so that acceptance must be unanimous), and whenever B is unanimously accepted at some wealth w, also A is unanimously accepted at wealth w, making A "(weakly) more acceptable." Similarly, income distribution A unanimously-rejection dominates income distribution B, if members are given the right to veto rejection (so that rejection must be unanimous), and whenever A is unanimously rejected at some wealth w, also B is unanimously rejected at wealth w, making B "(weakly) more rejected," or "(weakly) *less* acceptable." Quite surprisingly,[7] each of these two domination relations leads to a *complete* ranking of income distributions, i.e., to an (ordinal) measure of welfare. In particular, working with unanimous acceptance domination leads to the Rawlsian maximin criterion, and working with unanimous rejection domination leads to maximization of the geometric mean of the distribution (known in finance as the Kelly criterion). The following proposition states this result.

Proposition 1 *Let A and B be two random variables with finite support in R^+. Then:*

1. *A unanimously-acceptance dominates B if and only if $\min_{a_i \in A} a_i \geq \min_{b_i \in B} b_i$.*

2. *A unanimously-rejection dominates B if and only if $E_{a_i \in A}[\log(a_i)] \geq E_{b_i \in B}[\log(b_i)]$.*

The proof is relegated to the appendix. Part (1) of Proposition 1 invokes the Rawlsian maximin rule, because by giving all the members of society the right to veto acceptance when the extent of risk aversion is not bounded, we guarantee that for any wealth level $w > \min_{a_i \in A} a_i$, there would be some $u \in U^*$ that would be risk averse enough to reject income distribution A at w. This is a new presentation of the Rawlsian reasoning that emphasizes the extreme extent of risk-aversion embedded in it.

Part (2) of Proposition 1 invokes the usage of the geometric mean as a measure of welfare. Geometric means have some properties that make them appropriate as measures

[7] Though less surprising to those familiar with Hart (2011), from which the technic of unanimous-rejection was borrowed.

of welfare. First, the geometric mean has the feature that if all incomes in the economy are multiplied by some positive factor λ, the geometric mean is also multiplied by that factor λ. Second, like the arithmetic mean, it captures the central tendency or typical value of the incomes in the distribution, but unlike the arithmetic mean it decreases under mean-preserving spreads, thus capturing the bad impact of inequality on welfare. Indeed, starting from 2010 the United Nations Human Development Index is calculated using geometric mean instead of arithmetic mean.

Both measures of welfare, the minimal income and the geometric mean of incomes, are invariant to replication of the population.[8]

4 Discussion

The criterion of unanimous rejection implies via Proposition 1 that the *log utility* can be considered as representing the "preference of society." Indeed, Arrow (1971) advocates the usage of the log utility in economic models because he finds this utility function to be the most reasonable one from a descriptive point of view. Accordingly, Blanchard and Fischer (1989) report that intertemporal choices suggest a relative risk aversion coefficient close to 1, in line with log utility. Metrick (1995) reports a similar risk-attitude among participants in the 'Jeopardy!' TV game. However, the descriptive appeal of the log utility is somewhat controversial. In particular, some consider it to be too risk-loving to be descriptive of human behavior. Dasgupta (1998) and Friend and Blume (1975) present evidence that suggest a risk aversion coefficient around 2. Explaining the

[8]This is an important feature for a measure of welfare (see Shorrocks 1983 and Sen 1976).

'equity premium puzzle' with vN-M utilities requires even a larger relative risk aversion coefficient (see e.g. Mehra and Prescott 1985). Cohen and Einav (2007) shed light on a possible source for these seemingly contradictions. In a report on an experimental work that aims to measure risk attitudes, they get (for their benchmark model) that the *mean* relative risk aversion coefficient for their subjects was 97.22, while the *median* relative risk aversion coefficient for those same subjects was 0.37, i.e., smaller in orders of magnitude than the reported mean (ibid, Table 5). This huge discrepancy between the mean evaluation and the median evaluation raises some doubts about the validity of empirically and experimentally based evaluations of the 'typical' attitude toward risk of a representative agent. Note however that in our model the log utility is not chosen in order to represent a typical agent, but is rather the result of using the preferences of a large set of utilities that satisfy some desirable properties.

In this paper we mixed normative and positive considerations. On the one hand, we deal with the normative question of what constitutes higher welfare, and what should be considered as more desirable 'behind a veil of ignorance.' On the other hand, we construct the set of utilities while trying to reflect descriptive characteristics of human behavior, such as weakly increasing relative risk aversion. This approach results in a special role for the log utility (a result we believe to be non obvious). And though it is true that at least in some contexts the log utility is justifiably considered to be not enough risk averse to characterize human preferences, we believe that some of its properties, fore and foremost among them the rejection *at any wealth level* of *any distribution* whose support includes the zero outcome, make it quite risk-averse.

5 Appendix

Lemma 2 *Let $u_1, u_2 \in U^*$ be two utility functions with absolute risk aversion coefficients ρ_1 and ρ_2 respectively, where $\rho_1(w) \geq \rho_2(w)$ for every $w > 0$. Then for every income distribution D with finite support in R^+ and every $w > 0$, if u_2 rejects D at w then u_1 rejects D at w too.*[9]

Proof. Let ψ be such that $u_1 = \psi \circ u_2$; then ψ is strictly increasing (since u_1 and u_2 are such), and concave (since for every $w > 0$ we have $\psi\prime(u_2(w)) = u_1'(w)/u_2'(w)$, hence $(\log \psi\prime(u_2(w)))' = (\log u_1'(w))' - (\log u_2'(w))' = -\rho_1(w) + \rho_2(w) \leq 0$, and so $\psi'' \leq 0$). Therefore $E_D[u_2(d_i)] \leq u_2(w)$ implies by Jensen inequality and by the monotonicity of ψ that $E_D[u_1(d_i)] = E_D[\psi(u_2(d_i))] \leq \psi(E_D[u_2(d_i)]) \leq \psi(u_2(w)) = u_1(w)$. ■

Proof of Proposition 1

1. *A unanimously-acceptance dominates B if and only if $\min\limits_{a_i \in A} a_i \geq \min\limits_{b_i \in B} b_i$.*

Proof. For any income distribution D, let $L(D) \equiv \min\limits_{d_i \in D} d_i$ and let $M(D) \equiv \max\limits_{d_i \in D} d_i$. Notice that for any income distribution D, and every $u \in U^*$, u accepts D at any $w \leq L(D)$. Now let $w_0 = L(D) + \epsilon$ for some arbitrary small and positive ϵ, and let p_L denote the non-zero probability of $L(D) \in \text{supp}(D)$. Now let $\hat{u}_\alpha(w) := (\log(\alpha w) - 1)/e$ for $w \leq \frac{1}{\alpha}$ and $\hat{u}_\alpha(w) := -\exp(-\alpha w)$ for $w \geq \frac{1}{\alpha}$; then $\gamma_{\hat{u}_\alpha}(w) = 1$ for $w \leq \frac{1}{\alpha}$ and $\gamma_{\hat{u}_\alpha}(w) = \alpha w$ for $w \geq \frac{1}{\alpha}$ and so $\hat{u}_\alpha(w) \in U^*$ for each $\alpha > 0$.[10] For every $\alpha > \frac{1}{L(D)}$,

[9] *Lemma 2 and its proof and the proof to the second part of Proposition 1 borrow heavily from Hart (2011).*
[10] $\hat{u}_\alpha(w)$ was originally proposed by Sergiu Hart in Hart (2011).

we get that $\hat{u}_\alpha(w)$ rejects distribution D at w_0 if and only if $\hat{u}_\alpha(w_0) \geq E_D[\hat{u}_\alpha(d_i)]$. A sufficient condition for a rejection of D by $\hat{u}_\alpha(w)$ at w_0 is then

$$-\exp(-\alpha w_0) = -\exp(-\alpha(L(D)+\epsilon)) \geq p_L(-\exp(-\alpha L(D))) + (1-p_L)(-\exp(-\alpha M(D))).$$

Dividing by $-\exp(-\alpha L(D))$ and setting $\alpha(\epsilon) = -\frac{\log(p_L)}{\epsilon}$, we get that $\hat{u}_{\alpha(\epsilon)}(w)$ rejects distribution D at w_0, because $p_L \leq p_L + (1-p_L)\, p_L^{\frac{M(D)-L(D)}{\epsilon}}$. Letting $\epsilon \to 0$, we get that $\forall \epsilon > 0$, $\exists \alpha(\epsilon)$ such that every $\hat{u}_\alpha(w) \in U^*$ with $\alpha > \alpha(\epsilon)$ rejects D at wealth level $L(D)+\epsilon$. Consequentially, income distribution A is unanimously accepted at w if and only if $w \leq L(A)$, and income distribution B is unanimously accepted at w if and only if $w \leq L(B)$, and so A unanimously-acceptance dominates B if and only if $L(A) \geq L(B)$.

■

2. *A unanimously-rejection dominates B if and only if* $E_{a_i \in A}[\log(a_i)] > E_{b_i \in B}[\log(b_i)]$.

Proof. For any income distribution D and any wealth level $w > 0$, we will show that D is unanimously rejected at w if and only if $\log(w) \geq E_D[\log(d_i)]$. This means that unanimous rejection by all $u \in U^*$ implies and is implied by rejection by the log utility, hence unanimously-rejection domination boils down to the preferences of the log utility (i.e., A unanimously-rejection dominates B if and only if $E_{a_i \in A}[\log(a_i)] > E_{b_i \in B}[\log(b_i)]$). The "only if" direction is immediate: if $\log(w) < E_D[\log(d_i)]$, then $u_l \equiv \log(w) \in U^*$ accepts D at w, thus D is not unanimously rejected at w. The proof for the "if" direction is less straight forward. Lemma 8 in Hart (2011) implies that if u is increasing and strictly concave, and $\lim_{w \to 0^+} u(w) = -\infty$, then $\lim_{w \to 0^+} \gamma_u(w) \geq 1$. From property (3) of

U^* (weakly increasing $\gamma_u(w)$), it follows that $\forall u \in U^*$, $\gamma_u(w) \geq 1$ for all $w > 0$. Thus, Lemma 2 implies that $\forall u \in U^*$, if $u_2 = u_l$ rejects D at w then $u_1 = u$ rejects D at w too (because $\gamma_u(w) \geq 1 = \gamma_{u_l}(w) \Rightarrow \rho_u(w) \geq \rho_{u_l}(w)$), i.e., $\log(w) \geq E_D[\log(d_i)]$ implies that D is unanimously rejected at w. ■

6 References

References

[1] Arrow, Kenneth J. (1965), *Aspects of the Theory of Risk-Bearing*. Helsinki: Yrjö Jahnssonin Säätiö.

[2] ——— (1971), *Essays in the Theory of Risk Bearing*. Chicago: Markham.

[3] Blanchard O. and S. Fischer (1989), *Lectures on Macroeconomics*. Cambridge, Mass.: MIT Press.

[4] Carlsson, F., Daruvala, D., and O. Johansson-Stenman (2005), "Are People Inequality-Averse, or Just Risk-Averse?," *Economica* 72, 375–396.

[5] Cohen, A. and L. Einav (2007), "Estimating Risk Preferences from Deductible Choice," *American Economic Review,* 97(3), 745-788.

[6] Dasgupta, P. (1998), "Population, consumption and resources: ethical issues," *Ecological Economics* 24, 139–52.

[7] Ebert, U. (1987), "Size and distribution of incomes as determinants of social welfare," *Journal of Economic Theory*, 41, 23-33.

[8] Foster, D. P. and S. Hart (2009), "An Operational Measure of Riskiness," *Journal*

of Political Economy 117, 785–814.

[9] Friend, I. and M.E. Blume (1975), "The demand for risky assets," *American Economic Review* 65, 900-922.

[10] Harsanyi, J. C. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63(4), 309-321.

[11] Hart, S. (2011), "Comparing Risks by Acceptance and Rejection," *Journal of Political Economy* 119(4), 617-638.

[12] Mehra R. and E. C. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, 145-161.

[13] Metrick, A. (1995), "A Natural Experiment in 'Jeopardy!,"' *American Economic Review*, 85(1), 240–53.

[14] Rawls, J. (1971), *A Theory of Justice.* Cambridge, Mass.: Harvard University Press.

[15] Safra, Z. and U. Segal (1998), "Constant Risk Aversion," *journal of economic theory*, 83, 19-42.

[16] Sen, A. K. (1976), "Real national income," *Review of Economic Studies*, 43, 19-39.

[17] Sheshinski, E. (1972), "Relation between a social welfare function and the Gini index of income inequality," *Journal of Economic Theory*, 4(1), 98-100.

[18] Shorrocks, A. F. (1983), "Ranking income distributions," *Econometrica*, 50, 3-17.