KU LEUVEN

DEPARTMENT OF ECONOMICS

Interpersonal comparisons by means of money metric utilities: why one should use the same reference prices for all

Bart CAPÉAU, André DECOSTER and Liebrecht DE SADELEER

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS19.05

MARCH 2019

Interpersonal comparisons by means of money metric utilities: Why one should use the same reference prices for all

Bart Capéau, André Decoster, and Liebrecht De Sadeleer*

March 2019

Abstract

We show that using different reference prices for different individuals in money metrics of well–being leads to violations of several normative properties of interpersonal welfare comparisons that have become popular in the fairness literature. An empirical illustration for Belgian single adults available for the labour market in 2015 shows that the violation of these principles in the labour consumption context, is all but exceptional.

Key words: Money metric utility, well–being measurement, equivalent variation, labour supply.
JEL :codes: D61, D63, I31, J22.

1 Introduction

Money metric utilities (MMU's) measure individual well–being or welfare – both terms will be used interchangeably in this paper – by means of the expenditure function evaluated at a given vector of reference prices. They have become a standard tool in (applied) welfare analysis. They serve *e.g.* as a basis for calculating equivalent and compensating variations (see *e.g.* King, 1983, 1987, Blundell, Preston, and Walker, 1994, Aaberge *et al.*, 1995, or Creedy and Kalb, 2005). Recently there has been renewed attention to these measures by the fairness approach in social choice theory (Fleurbaey, 2008, 2011, Fleurbaey and Blanchet, 2013, Fleurbaey and Maniquet, 2017, 2018).

For practical applications, the question arises as to which reference prices to use. In the context of labour consumption problems, a common practice is to use actual wages as the reference price for leisure (*e.g.* Aaberge *et al.*, 1995 and Creedy and Kalb, 2005). This choice has recently also been advocated for by Chiappori (2015) and Chiappori and Meghir (2016), and is used in many

* Department of Economics, KU Leuven. Corresponding author: Bart Capéau. E–mail: bart.capeau@kuleuven.be.
We would like to thank participants to the Modelling Tax Policy and Compliance workshop at TARC – Exeter Business School March 14–15 2019 for valuable comments. Bart Capéau and André Decoster gratefully acknowledge financial support from BELSPO through grants BR/121/A5/MEqIn and BR/132/A4/BEL-Ageing. This paper uses data from the national EU–SILC (Statistics on Income and Living Conditions) survey for Belgium, made available by Statistics Belgium.

applications of the so–called collective household model (*e.g.* Cherchye *et al.*, 2015, 2018, or Ferrando, 2018). Given individual wage heterogeneity, the consequence is that different reference prices are used for evaluating each person's welfare.

In this paper, we show that such a practice leads to the violation of several normative principles for interpersonal welfare comparisons from fair social choice theory (Fleurbaey and Maniquet, 2011, Decancq, Fleurbaey, and Schokkaert, 2015). More precisely, we show that in such a case the Same Preference Principle and the Nested Contour Criterion are violated. Moreover, in such a case, these measures lose the Nested Opportunity Set Property, which was one of the reasons why these measures had renewed attention from the fairness approach to social choice (Fleurbaey, 2011). Furthermore, we investigate some implications of the Same Preference Principle and the Nested Contour Criterion for the measurement of welfare differences. We show that the well–known candidates for measuring such welfare differences, the equivalent and compensating variations, violate these implied properties.

Our contribution does not help to choose the appropriate vector of reference prices. It is well known that this choice does matter for making interpersonal welfare comparisons (see Chiappori and Meghir, 2016 for an example). In order to try to circumvent this problem, one strand in the literature has therefore focussed on investigating which class of preferences allows money metric utilities to be reference price independent, *i.e.* the choice of reference prices does not affect welfare rankings of individuals (see *e.g.* Blackorby *et al.* 1994). This turns out to be overly restrictive (homotheticity of preferences). We think however that the choice of reference prices should follow from normative principles guiding interpersonal welfare comparisons, deemed to prevail irrespective of the particular shape of individual preferences to which they are applied, and therefore reference price independence is not necessarily an attractive property.[1]

Our contribution is related to Pauwels (1978, 1986), who alerts us to potential incoherences when using the compensating variation, which is essentially the result of applying a money metric utility when making intra–individual welfare comparisons. Pauwels (1978, 1986) shows that if the welfare cost according to the compensating variation for an individual switching from the baseline to situation 1 is higher than that of switching to situation 2, it does not necessarily follow that situation 2 is better than situation 1. The reason is that with the compensating variation, the welfare change of the switch to situation 1 is measured at prices of situation 1, while that of a switch to situation 2 is measured at prices in situation 2. If we apply the common practice of using market wages to construct money metric utilities, and we compare an individual in two situations in which she obtains different wages, a similar problem arises. We will show that such a practice is not consistent with a person's own ranking of these situations according to her preferences. In the jargon of the fairness literature on well–being measurement this is called a violation of the principle of Respect of Individual Preferences. We extend that critique to the case of interpersonal comparisons. If two individuals with the same preferences face *e.g.* different wages, the well–being ranking of those persons resulting from using the wage of each as a reference price for evaluating the

[1] Potential candidates of such principles can be found in Capéau *et al.* (2016).

money metric would not necessarily follow these persons' own ranking according to their identical preferences. The measure would violate the Same Preference Principle, in the terminology of the fairness literature.

The next section sets out the model and contains a formal definition of the class of money metric well–being measures (Subsection 2.1) and some normative principles of well–being measurement (Subsection 2.2). Section 3 shows that money metrics with different reference prices across individuals fail to satisfy all but one of the principles from Subsection 2.2. Section 4 gives the model in the specific case of labour and consumption as the only determinants of well–being, and shows how in this particular context the MMU's on which compensating and equivalent variation concepts of measuring the welfare cost of tax reforms are based, suffer from the flaws exhibited in Section 3. In Section 5, we show that the equivalent variation itself does not satisfy the requirements on the measurement of welfare differences implied by the Same Preference principle and the Nested Contour Criterion. Section 6 contains an empirical application. Section 7 concludes.

2 Notation and definitions

Individuals are indexed by $i, j, k, \ldots$. They have preferences, R_i, over commodity bundles, denoted by $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$.[2] One of the goods can be leisure. Commodity bundles belong to the consumption set X, which is a non–empty comprehensive subset of $\mathbb{R}^n_+$. Comprehensiveness means that when $\mathbf{x} \in X$, then all $\mathbf{y} \in \mathbb{R}^n_+$ such that $\mathbf{y} \leq \mathbf{x}$, belong to X as well.[3] So, the origin $\mathbf{0}$ is part of the consumption set. The power set of X, that is the set of all subsets of X, will be denoted by $\mathcal{X}$. We follow the usual notation for subsets: $A \subseteq B$ denotes that A is a subset of (potentially equal to) B, while $A \subset B$ means that A is a strict subset of B. X is endowed with the relative topology induced by the natural (Euclidean) topology on $\mathbb{R}^n$.

Preferences are complete, transitive, and continuous binary relations on X: $\mathbf{x} \, R_i \, \mathbf{y}$ means that i weakly prefers bundle $\mathbf{x}$ over $\mathbf{y}$.[4] It is furthermore assumed that $\mathbf{x} \, R \, \mathbf{0}$ for all $\mathbf{x} \in X$. The domain of such preferences is denoted by $\mathcal{R}$. Indifference and strict preference will be denoted respectively by I_i and P_i.

Definition 1 (**Upper and Lower Contour Set, and Indifference Set**). *The upper contour set of a bundle* $\mathbf{x}$ *according to preference ordering* R*, say* $UC(\mathbf{x}, R)$*, is the set of all bundles weakly preferred to* $\mathbf{x}$ *according to* R*:*

$$UC(\mathbf{x}, R) = \{\mathbf{y} \in X \,|\mathbf{y} \, R \, \mathbf{x}\}. \tag{1}$$

[2] We drop the subscript i of the preference ordering R_i, if the reference to a specific person is not necessary.

[3] For vector inequalities, $\mathbf{x} \leq \mathbf{y}$ means that all coordinates of $\mathbf{y}$ are larger than or equal to the corresponding coordinates of $\mathbf{x}$.

[4] A preference ordering is continuous if its lower and upper contour sets are closed. The definitions of lower and upper contour sets are given below (see Definition 1).

The lower contour set of a bundle $\mathbf{x}$ *according to preference ordering* R*, say* $LC(\mathbf{x}, R)$*, is the set of bundles that are weakly worse than* $\mathbf{x}$ *according to* R*:*

$$LC(\mathbf{x}, R) = \{\mathbf{y} \in X \,|\mathbf{x}\, R\, \mathbf{y}\}. \tag{2}$$

The indifference set, say $IC(\mathbf{x}, R)$*, is the set of bundles which are equally as good as* $\mathbf{x}$ *according to* R*, and it thus coincides with the intersection of the lower and upper contour set:*

$$IC(\mathbf{x}, R) = \{\mathbf{y} \in X \,|\mathbf{y}\, I\, \mathbf{x}\} = LC(\mathbf{x}, R) \cap UC(\mathbf{x}, R). \tag{3}$$

2.1 Money metric utilities

In general, individual well–being measurement serves to compare the well–being of different individuals in a given situation (that is, to determine who is better off than who), or of a given individual in different situations. In line with the fairness approach to well–being measurement (Fleurbaey and Maniquet, 2018), we assume that individual well–being does not only depend on the bundle of goods obtained, say $\mathbf{x}$, but also on an individual's preferences.[5] Individual welfare measurement thus engenders a binary relation on pairs of commodity bundles and preference orderings. To motivate this, consider the following example. A full time job with high income might be preferred to unemployment and low replacement income for someone with relatively low preference intensity for leisure, while the reverse may be the case for someone less eager to work. Adding this alternative to the choice set of both individuals, therefore, has not the same welfare implications.

We assume that the binary well–being relation on pairs of commodity bundles and preferences, can be represented by a well–being measure or metric, say $W_i(\mathbf{x}, R)$, with $W_i(\mathbf{x}, R_i) \geq W_j(\mathbf{y}, R_j)$ meaning that an individual i with preferences R_i and commodity bundle $\mathbf{x}$ is at least as well off as an individual j with preferences R_j and commodity bundle $\mathbf{y}$. We implicitly allow for the welfare measure to depend on other information than the bundle $\mathbf{x}$, and preferences R, by indexing W on i.

The purpose of the present paper is to highlight some properties of the class of money metric utilities, when used for the purpose of individual well–being measurement. The elements of this class of utility measures are indexed by a strictly positive reference price vector, $\mathbf{p}^r$, and formally defined as follows.

Definition 2 (**Money Metric Utility**, MMU)**.** *Given a reference price vector* $\mathbf{p}^r \in \mathbb{R}^n_{++}$*, the money metric utility* $MMU(\mathbf{x}, R; \mathbf{p}^r)$*, is a mapping which associates with each pair of a commodity bundle* $\mathbf{x}$ *and a preference ordering* R*, the minimal budget necessary to buy a bundle at reference prices* $\mathbf{p}^r$*, which is at least as good as* $\mathbf{x}$ *according to preferences* R*:*

$$MMU(\cdot; \mathbf{p}^r) : X \times \mathcal{R} \to \mathbb{R} : (\mathbf{x}, R) \mapsto MMU(\mathbf{x}, R; \mathbf{p}^r) = \inf\left\{\mathbf{p}^{r\prime}\mathbf{z} \,|\mathbf{z} \in X \text{ and } \mathbf{z}\, R\, \mathbf{x}\right\}. \tag{4}$$

[5] By preferences, we mean here considered judgements of a person of different situations according to their conception of a good life. We thus neglect the questions arising from behavioural economics on the existence of such judgements (see *e.g.* Gul and Pesendorfer, 2001).

The reader may verify that when using a functional representation of R by a utility function, say $U = u(\mathbf{x})$, the money metric utility coincides with the expenditure function $e(\mathbf{p}^r; U)$ as a representation of preferences.[6] Given the assumption that preferences are continuous, the infimum in Equation (4) will be the minimum.

By using money metric utilities for measuring individual well–being, we mean that we select for each individual i an element from the class of money metric utility functions by fixing a reference price $\mathbf{p}_i^r$, such that:

$$W_i(\mathbf{x}, R_i) = MMU(\mathbf{x}, R_i; \mathbf{p}_i^r). \tag{5}$$

2.2 Principles of fair well–being measurement

We now list some principles of individual well–being measurement. The first one says that an individual well–being measure should not go against an individual's own judgement concerning his/her situation, according to his/her own preferences.

Definition 3 (**Respect for Individual Preferences**, RIP (Decancq, Fleurbaey, and Schokkaert, 2015))**.** *An individual well–being measure $W_i(\mathbf{x}, R)$ satisfies* Respect for Individual Preferences (RIP)*, if for any individual, i, for any preference ordering, R, and for any pair of bundles $\mathbf{x}$ and $\mathbf{y}$, it holds that:*

$$W_i(\mathbf{x}, R) \geq W_i(\mathbf{y}, R) \quad \Leftrightarrow \quad \mathbf{x}\, R\, \mathbf{y}. \tag{6}$$

The next principle extends RIP to cases of different persons, but with the same preferences. That is, when comparing individuals with the same preferences, these preferences completely determine how these individuals will be mutually compared if this principle is to be respected.

Definition 4 (**Same Preference Principle**, SPP (Decancq, Fleurbaey, and Schokkaert, 2015))**.** *An individual well–being measure $W_i(\mathbf{x}, R)$ satisfies the* Same Preference Principle (SPP)*, if for any pair of individuals, i and j, with the same preferences, whatever they may be, $R_j = R_i \equiv R$ say, and for any pair of bundles $\mathbf{x}$ and $\mathbf{y}$, it holds that:*

$$W_i(\mathbf{x}, R) \geq W_j(\mathbf{y}, R) \quad \Leftrightarrow \quad \mathbf{x}\, R\, \mathbf{y}. \tag{7}$$

Consider now the case where two individuals do not exhibit the same preferences. Assume a situation where all bundles that are at least as good as the one actually obtained by one of them according to his own preferences, are strictly better than all goods that are not better than the one actually obtained by the other according to her preferences. Then, it might be reasonable to conclude that the former person is better off than the latter. This property is stated formally in the next definition.

[6] See Weymark (1985) for technical conditions under which MMU's can serve as (continuous) representations of preferences.

Definition 5 (**Nested Contours Criterion**, NCC (Decancq, Fleurbaey, and Maniquet (2015))**.** *An individual well–being measure* $W_i(\mathbf{x}, R)$ *satisfies the* Nested Contour Criterion (NCC)*, if for any pair of individuals, i and j, and for any pair of bundles* $\mathbf{x}$ *and* $\mathbf{y}$*, it holds that:*

$$LC(\mathbf{y}, R_j) \cap UC(\mathbf{x}, R_i) = \varnothing \quad \Rightarrow \quad W_i(\mathbf{x}, R_i) > W_j(\mathbf{y}, R_j). \tag{8}$$

The equal opportunity literature (Fleurbaey, 2008, Roemer and Trannoy, 2015, 2016) has advocated to measure well–being in terms of available opportunities rather than actual welfare obtained by choosing one alternative from the available opportunities, as the latter is a matter of choice for which individuals bear sole responsibility. It has been argued however that comparing opportunity sets without taking into account preferences is not very attractive (Bossert, Pattanaik, and Xu, 1994). Indeed, alleviating income taxes on high wages may allow the one with less intense preferences for leisure to improve while this does not necessarily hold for someone more intensely preferring leisure. One way to measure welfare in terms of opportunity sets, while taking into account the extent to which such additional opportunities potentially improve welfare is imposing the following property: a person has higher welfare than someone else if it is possible to attribute to her/him an opportunity set such that the best she/he can reach according to her/his preferences given these opportunities is equally as good as the actual situation for which welfare has to be measured, and this set includes a similarly defined opportunity set for the other person. Formally, this reads as follows.

Definition 6 (**Nested Opportunity Set Property**, NOSP)**.** *An individual well–being measure* $W_i(\mathbf{x}, R)$ *satisfies the* Nested Opportunity Set Property (NOSP)*, if there exists a mapping from pairs of commodity bundles and preferences to subsets of the commodity space,* $B : X \times \mathcal{R} \to \mathcal{X} : (\mathbf{x}, R) \mapsto B(\mathbf{x}, R)$*, such that for any pair of individuals, i and j, any pair of preference orderings* R_i *and* R_j*, and for any pair of bundles* $\mathbf{x}$ *and* $\mathbf{y}$*:*

$$\begin{aligned} W_i(\mathbf{x}, R_i) \geq W_j(\mathbf{y}, R_j) \quad \Leftrightarrow \quad & B(\mathbf{x}, R_i) \supseteq B(\mathbf{y}, R_j), \\ & \forall \mathbf{z} \in \arg\max_{\mathbf{w}} \{R_i \,|\mathbf{w} \in B(\mathbf{x}, R_i)\} : \ \mathbf{x} I_i \mathbf{z}, \ \textit{and} \\ & \forall \mathbf{z} \in \arg\max_{\mathbf{w}} \{R_j \,|\mathbf{w} \in B(\mathbf{y}, R_j)\} : \ \mathbf{y} I_j \mathbf{z} \end{aligned} \tag{9}$$

.

If all individuals would have equal preferences, say R, this property can always be satisfied by defining $B(\mathbf{x}, R)$ as follows:

$$B(\mathbf{x}, R) = LC(\mathbf{x}, R). \tag{10}$$

Even if the NCC is a criterion focussing on respecting preferences while NOSP focusses on opportunities, the latter criterion implies the former as is stated in the following claim.

Proposition 1. NOSP *implies* NCC.

Proof. Suppose NCC is not satisfied. That is, there exist $\mathbf{x}, \mathbf{y} \in X$, $R_i, R_j \in \mathcal{R}$, such that $W_j(\mathbf{y}, R_j) \geq W_i(\mathbf{x}, R_i)$ and $LC(\mathbf{y}, R_j) \cap UC(\mathbf{x}, R_i) = \varnothing$. Now $LC(\mathbf{y}, R_j) \cap UC(\mathbf{x}, R_i) = \varnothing \implies LC(\mathbf{y}, R_j) \subseteq LC(\mathbf{x}, R_i) \setminus IC(\mathbf{x}, R_i)$.

From NOSP it follows that $B(\mathbf{y}, R_j) \subseteq LC(\mathbf{y}, R_j)$, and if $W_j(\mathbf{y}, R_j) \geq W_i(\mathbf{x}, R_i)$ then $B(\mathbf{x}, R_i) \subseteq B(\mathbf{y}, R_j)$. So, $B(\mathbf{x}, R_i) \subseteq LC(\mathbf{x}, R_i) \setminus IC(\mathbf{x}, R_i)$. But NOSP imposes that $\exists \mathbf{z} \in B(\mathbf{x}, R_i) \cap IC(\mathbf{x}, R_i)$, a contradiction.

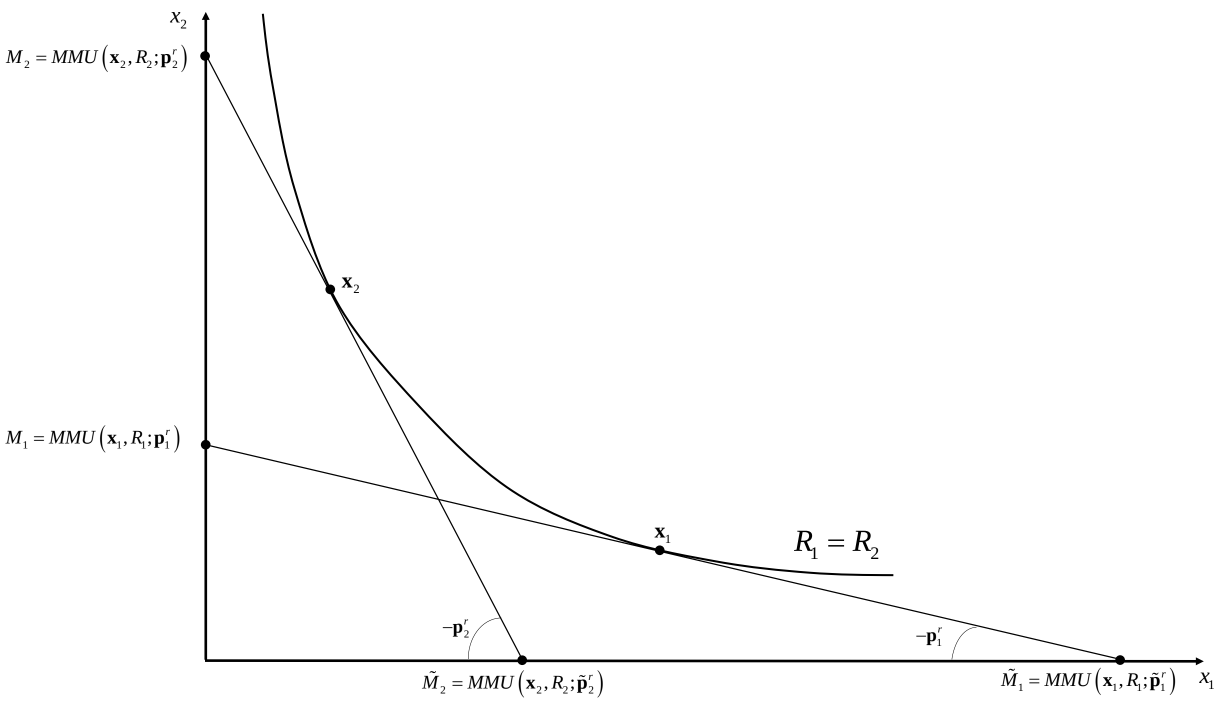
■

3 Results

In this section we evaluate some consequences of using $MMU(\mathbf{x}, R_i; \mathbf{p}_i^r)$ as a welfare measure $W_i(\mathbf{x}, R_i)$.

Result 1. *When using different reference prices for each individual, say $\mathbf{p}_1^r \neq \mathbf{p}_2^r$, the money metric utility $MMU(\mathbf{x}, R_i; \mathbf{p}_i^r)$ does not satisfy the Same Preference Principle* (SPP).

The validity of this claim is illustrated in Figure 1. In the figure we consider a case where two agents with the same preferences obtain respectively the bundle $\mathbf{x}_1$ and $\mathbf{x}_2$, on the same indifference curve: $\mathbf{x}_1 I_{1,2}\, \mathbf{x}_2$. For the sake of simplicity, and without loss of generality, we have chosen the reference prices such that $\mathbf{x}_i = \arg\max_{\mathbf{z}} \{R_i \,|\mathbf{p}_i^{r\prime}\mathbf{z} \leq \mathbf{p}_i^{r\prime}\mathbf{x}_i \text{ and } \mathbf{z} \in X\}$, for $i = 1, 2$. The same preference principle would then impose that both individuals obtain equal welfare, $W_1(\mathbf{x}_1, R_1) = W_2(\mathbf{x}_2, R_2)$, while $MMU(\mathbf{x}_1, R_1; \mathbf{p}_1^r) \neq MMU(\mathbf{x}_2, R_2; \mathbf{p}_2^r)$.



Note: in case of $\mathbf{p}_i^r$, the price of good 2 is normalised to one, while in $\tilde{\mathbf{p}}_i^r$, the price of good 1 is normalised to one.

Figure 1: MMU with different reference prices does not satisfy the Same Preference Principle.

A rather awkward consequence of using different reference prices for different persons is that price normalisation does matter. As illustrated in the figure, if we normalise prices such that the price of good 2 is the *numéraire* then $MMU(\mathbf{x}_2, R_2; \mathbf{p}_2^r) > MMU(\mathbf{x}_1, R_1; \mathbf{p}_1^r)$, and the reverse would hold when prices were normalised such that the price of good 1 equals one (denoted by $\tilde{\mathbf{p}}_i^r$, $i = 1, 2$ in the figure). One could of course try to find a normalisation such that $MMU(\mathbf{x}_2, R_2; \mathbf{p}_2^{rs}) = MMU(\mathbf{x}_1, R_1; \mathbf{p}_1^{rs})$, but this would yield very cumbersome normalisation rules. Moreover, the issue would reappear if more than two such individuals, each with their own reference prices, should be compared in terms of welfare.

Suppose now that the points $\mathbf{x}_1$ and $\mathbf{x}_2$ in Figure 1 represent two observations of the same individual at different points in time, facing different market prices for commodities, which have been used as respective reference prices for each of the observations. While the agent is indifferent between both situations, the respective money metric utilities give different values for both situations.[7] Hence, RIP is violated. However, when doing so, in fact, different money metric utilities are used to evaluate different points. Strictly speaking, this is not any more a money metric utility evaluation.[8] Inconsistency problems arising from such a practice were revealed amongst others by Boadway (1974), Pauwels (1978, 1986), King (1983, 1987), Blackorby and Donaldson (1990), and Donaldson (1992). Weymark (1985) considers the conditions under which a money metric utility is a (continuous) representation of preferences. As we will see in Proposition 2, these conditions are satisfied in our model. Therefore RIP is satisfied when using one particular money metric utility for each individual, be it a different one for each individual, *i.e.* using a different reference price for each individual.

Result 2. *When using different reference prices for each individual, say $\mathbf{p}_1^r \neq \mathbf{p}_2^r$, the money metric utility $MMU(\mathbf{x}, R_i; \mathbf{p}_i^r)$ does not satisfy the Nested Contours Criterion* (NCC).

Again, we illustrate the claim graphically (Figure 2). As $UC(\mathbf{x}_2, R_2) \cap LC(\mathbf{x}_1, R_1) = \varnothing$, we should have $W_2(\mathbf{x}_2, R_2) > W_1(\mathbf{x}_1, R_1)$, while $MMU(\mathbf{x}_1, R_1; \mathbf{p}_1^r) > MMU(\mathbf{x}_2, R_2; \mathbf{p}_2^r)$.

The figure also illustrates our next claim which says that when using different reference prices for each individual, the money metric utility does not satisfy NOSP. Suppose it does. As we have $MMU(\mathbf{x}_1, R_1; \mathbf{p}_1^r) > MMU(\mathbf{x}_2, R_2; \mathbf{p}_2^r)$, it should be possible to find subsets $B(\mathbf{x}_1, R_1)$ and $B(\mathbf{x}_2, R_2)$ of X such that $B(\mathbf{x}_2, R_2) \subset B(\mathbf{x}_1, R_1) \subseteq LC(\mathbf{x}_1, R_1)$, and, in addition, there exists at least some $\mathbf{z} \in B(\mathbf{x}_2, R_2)$ such that $\mathbf{z} I_2 \mathbf{x}_2$. This is not possible as $LC(\mathbf{x}_1, R_1) \cap IC(\mathbf{x}_2, R_2) = \varnothing$. This confirms Proposition 1 stating that NOSP is a strengthening of NCC, or stated differently, a violation of NCC implies a violation of NOSP.

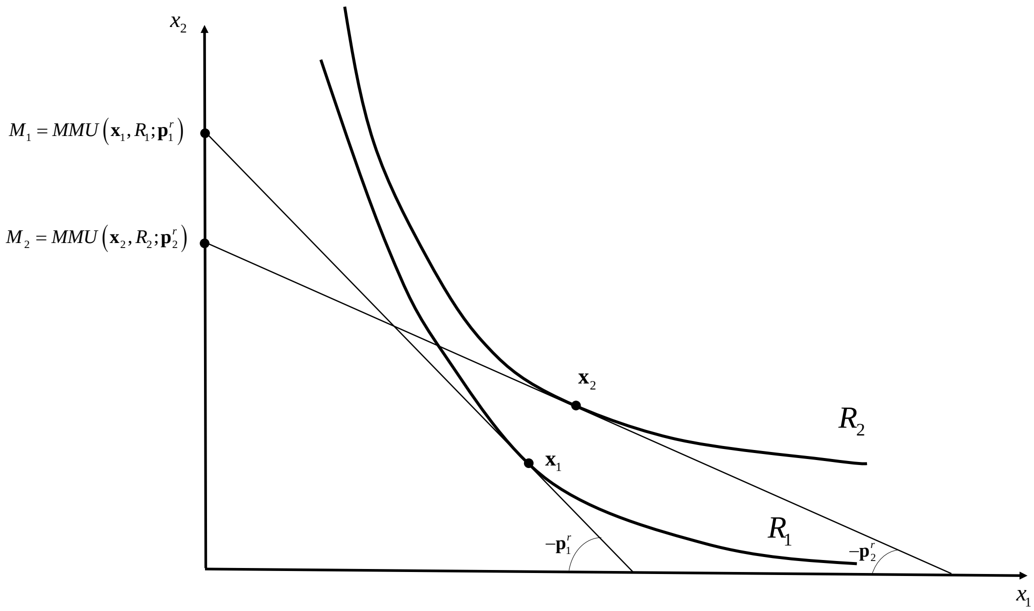
We conclude by stating this claim formally.

Result 3. *When using different reference prices for each individual, say $\mathbf{p}_1^r \neq \mathbf{p}_2^r$, the money metric utility $MMU(\mathbf{x}, R_i; \mathbf{p}_i^r)$ does not satisfy the Nested Opportunity Set Property* (NOSP).

None of the problems caused by using different reference prices across agents mentioned in Results 1–

[7] In general, for such a case to occur, income will need to adapt too.

[8] Recall that according to the principle used to construct an individual well–being measure from money metric utilities in Equation (4), the reference price for an individual i should not vary over commodity bundles.



Note: the price of good 2 is normalised to one.

Figure 2: MMU with different reference prices does not satisfy the Nested Contours Criterion.

3, would occur if we chose one and the same reference price vector for all respondents. This is formally stated in the next proposition.

Proposition 2. *The* MMU *with common reference prices for all agents satisfies* RIP, SPP, NCC, *and* NOSP.

Proof. Before proving the statement itself, we define some concepts that will operate as tools in the proofs of RIP, SPP, NCC, and NOSP and prove some of their properties.

First, as the infimum in the definition of the MMU is attained in case preferences are continuous, we can define $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$ as a commodity bundle such that $MMU(\mathbf{x}, R; \mathbf{p}^r) = \mathbf{p}^{r\prime}\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$ and $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)\, R\, \mathbf{x}$. If there is more than one such a bundle, it does not matter which one is chosen to be $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$.

Next, define $A(m, \mathbf{p}^r)$ as the budget set corresponding to a budget m with price vector $\mathbf{p}^r$, that is:

$$A(m, \mathbf{p}^r) := \{\mathbf{z} \in X \,|\mathbf{p}^{r\prime}\mathbf{z} \leq m\}\,. \tag{11}$$

Hence, this function A evaluated at a money metric utility $MMU(\mathbf{x}, R; \mathbf{p}^r)$ and price vector $\mathbf{p}^r$, is the budget set corresponding to the money metric utility with reference price vector $\mathbf{p}^r$ of a given pair of a commodity bundle $\mathbf{x}$ and a preference ranking R. Note also that $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r) \in A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r)$.

Moreover, let $m^*(\mathbf{p}^r) := \sup\{\mathbf{p}^{r\prime}\mathbf{x} | \mathbf{x} \in X\}$ where the supremum is taken in $\mathbb{R}_+ \cup \{\infty\}$. Then, by definition, it follows that for all $\mathbf{x} \in X$ and $m > m^*(\mathbf{p}^r)$: $\mathbf{p}^{r\prime}\mathbf{x} < m$. Furthermore, for every $(\mathbf{x}, R) \in X \times \mathcal{R}$ and every reference price vector $\mathbf{p}^r$, it holds that $MMU(\mathbf{x}, R; \mathbf{p}^r) = \mathbf{p}^{r\prime}\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r) \leq m^*(\mathbf{p}^r)$.

As X is comprehensive and all prices in $\mathbf{p}^r$ are strictly positive, the set $\{A(m, \mathbf{p}^r) | \, m \in [0, m^*]\}$ is ordered with respect to inclusion:

$$\forall \, m, m' \in [0, m^*] : \; A(m, \mathbf{p}^r) \subseteq A(m', \mathbf{p}^r) \Leftrightarrow m \leq m'. \tag{12}$$

We now show by contradiction that

$$\forall \, (\mathbf{x}, R) \in X \times \mathcal{R} : \; A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r) \subseteq LC(\mathbf{x}, R). \tag{13}$$

Suppose the claim does not hold. That is, there exists a $\mathbf{z} \in A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r)$ such that $\mathbf{z} \, P \, \mathbf{x}$. By our assumption on preferences, $\mathbf{y} \, R \, \mathbf{0}$ for all $\mathbf{y} \in X$ and all $R \in \mathcal{R}$. Hence $\mathbf{z} \neq \mathbf{0}$. As all elements in $\mathbf{p}^r$ are strictly positive, this implies that $\mathbf{p}^{r\prime}\mathbf{z} > 0$. Continuity of preferences implies that the strict better than set of $\mathbf{x}$ (that is $UC(\mathbf{x}, R) \setminus LC(\mathbf{x}, R)$) is open. So, there exists a small open neighbourhood of $\mathbf{z}$, say $N_\epsilon(\mathbf{z})$, such that $\mathbf{y} \, P \, \mathbf{x}$ for all $\mathbf{y} \in N_\epsilon(\mathbf{z})$. By comprehensiveness of X and strict positiveness of prices, any small open neighbourhood of $\mathbf{z}$ contains points $\mathbf{w}$ such that $\mathbf{p}^{r\prime}\mathbf{w} < \mathbf{p}^{r\prime}\mathbf{z}$. Indeed, $N_\epsilon(\mathbf{z})$ contains an element $\mathbf{w} \in X$ such that $\mathbf{w} \leq \mathbf{z}$ and $\mathbf{w} \neq \mathbf{z}$. By definition of A, $\mathbf{p}^{r\prime}\mathbf{z} \leq MMU(\mathbf{x}, R; \mathbf{p}^r)$. But then, $\mathbf{w} \, P \, \mathbf{x}$ and $\mathbf{p}^{r\prime}\mathbf{w} < MMU(\mathbf{x}, R; \mathbf{p}^r)$, contradicting the definition of MMU.

We are now ready to prove the properties RIP, SPP, NCC, and NOSP.

- RIP and SPP: SPP is a strengthening of RIP, so we only need to prove that $MMU(\cdot; \mathbf{p}^r)$ satisfies SPP.

  Sufficiency: $MMU(\mathbf{x}, R; \mathbf{p}^r) \geq MMU(\mathbf{y}, R; \mathbf{p}^r) \Rightarrow \mathbf{x} \, R \, \mathbf{y}$.

  As $\{A(m, \mathbf{p}^r) | \, m \in [0, m^*]\}$ is ordered with respect to inclusion, it follows that $A(MMU(\mathbf{y}, R; \mathbf{p}^r), \mathbf{p}^r) \subseteq A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r)$. Furthermore, we have established in Equation (13) that $A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r) \subseteq LC(\mathbf{x}, R)$. So, $A(MMU(\mathbf{y}, R; \mathbf{p}^r), \mathbf{p}^r) \subseteq A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r) \subseteq LC(\mathbf{x}, R)$.

  As $\mathbf{z}^*(\mathbf{y}, R; \mathbf{p}^r) \in A(MMU(\mathbf{y}, R; \mathbf{p}^r), \mathbf{p}^r)$, $\mathbf{z}^*(\mathbf{y}, R; \mathbf{p}^r) \in LC(\mathbf{x}, R)$. Thus, $\mathbf{x} \, R \, \mathbf{z}^*(\mathbf{y}, R; \mathbf{p}^r)$. By definition, $\mathbf{z}^*(\mathbf{y}, R; \mathbf{p}^r) \, R \, \mathbf{y}$. So, by transitivity of R, it follows from the last two statements that $\mathbf{x} \, R \, \mathbf{y}$.

  Necessity: $\mathbf{x} \, R \, \mathbf{y} \Rightarrow MMU(\mathbf{x}, R; \mathbf{p}^r) \geq MMU(\mathbf{y}, R; \mathbf{p}^r)$.

  Suppose now that $\mathbf{x} \, R \, \mathbf{y}$. By definition of $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$, $MMU(\mathbf{x}, R; \mathbf{p}^r) = \mathbf{p}^{r\prime}\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$ and $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r) \, R \, \mathbf{x}$. By transitivity of R, it follows that $\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r) \, R \, \mathbf{y}$. By definition of MMU, it must then hold that $MMU(\mathbf{y}, R; \mathbf{p}^r) \leq \mathbf{p}^{r\prime}\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$. As $MMU(\mathbf{x}, R; \mathbf{p}^r) = \mathbf{p}^{r\prime}\mathbf{z}^*(\mathbf{x}, R; \mathbf{p}^r)$, $MMU(\mathbf{y}, R; \mathbf{p}^r) \leq MMU(\mathbf{x}, R; \mathbf{p}^r)$.

- NCC: Suppose NCC is not satisfied, then there must exist at least two commodity bundle preference pairs, say $(\mathbf{x}, R_i)$ and $(\mathbf{y}, R_j)$, such that $UC(\mathbf{y}, R_j) \cap LC(\mathbf{x}, R_i) = \varnothing$ and $MMU(\mathbf{x}, R_i; \mathbf{p}^r) \geq MMU(\mathbf{y}, R_j; \mathbf{p}^r)$. As $\{A(m, \mathbf{p}^r) | m \in [0, m^*]\}$ is ranked with respect to inclusion, we have $A(MMU(\mathbf{y}, R_j; \mathbf{p}^r), \mathbf{p}^r) \subseteq A(MMU(\mathbf{x}, R_i; \mathbf{p}^r), \mathbf{p}^r)$, and we have established in Equation (13) that $A(MMU(\mathbf{x}, R_i; \mathbf{p}^r), \mathbf{p}^r) \subseteq LC(\mathbf{x}, R_i)$. Hence, $A(MMU(\mathbf{y}, R_j; \mathbf{p}^r), \mathbf{p}^r) \subseteq LC(\mathbf{x}, R_i)$. However, we know that $\mathbf{z}^*(\mathbf{y}, R_j; \mathbf{p}^r) \in A(MMU(\mathbf{y}, R_j; \mathbf{p}^r), \mathbf{p}^r)$ and $\mathbf{z}^*(\mathbf{y}, R_j; \mathbf{p}^r) \, R_j \, \mathbf{y}$, hence $\mathbf{z}^*(\mathbf{y}, R_j; \mathbf{p}^r) \in UC(\mathbf{y}, R_j)$, which is impossible as $\mathbf{z}^*(\mathbf{y}, R_j; \mathbf{p}^r) \in LC(\mathbf{x}, R_i)$ and $UC(\mathbf{y}, R_j) \cap LC(\mathbf{x}, R_i) = \varnothing$.
- NOSP: Let $B(\mathbf{x}, R) = A(MMU(\mathbf{x}, R; \mathbf{p}^r), \mathbf{p}^r)$.

■

Notice that Proposition 1 indicates that the last part of the proof also establishes that $MMU(\cdot; \mathbf{p}^r)$ satisfies NCC. We preferred to give an independent proof of that statement however.

If we strengthened NCC by requiring that $LC(\mathbf{y}, R_j) \subseteq LC(\mathbf{x}, R_i) \Rightarrow W_i(\mathbf{x}, R_i) \geq W_j(\mathbf{y}, R_j)$, with strict inequality on the right hand side if $LC(\mathbf{y}, R_j) \subset LC(\mathbf{x}, R_i)$, then MMU with common reference prices would no longer satisfy this criterion. The surface metric proposed by De Sadeleer (2017), which evaluates the welfare of the consumer by measuring the n–dimensional volume of the lower contour set of an alternative over a specified range, does satisfy this strengthened NCC criterion. But it does not satisfy NOSP. There seems therefore to be a trade–off between NOSP and this strengthened version of NCC.

4 The labour consumption context

Money metrics are frequently used for evaluating individual welfare of bundles of disposable income, c, and labour time, h (Aaberge, Dagsvik, and Strøm, 1995, Aaberge, Colombino, Strøm, 2000, Creedy and Kalb, 2005, Dagsvik and Karlström, 2005, Dagsvik, Locatelli, and Strøm 2009). In a context with labour time, two versions of these metrics are available, the unearned income version and the full income version. According to these measures, individual welfare of a certain consumption and labour time bundle, (h, c), can be measured as the minimal unearned or full income needed in order to be indifferent between this bundle and the best bundle that can be obtained with that income in a counterfactual situation where one could earn w^{ref} per unit of labour supply. Total time endowment is denoted by T. So, the consumption set X equals $[0, T] \times \mathbb{R}_+$.

Definition 7. *Given an alternative (h, c), a reference wage w^{ref}, and a preference relation R, the unearned income money metric utility, say $MMU\left((h, c), R; w^{ref},\right)$, is the minimal unearned income level m such that there still exists a bundle $\left(h', m + w^{ref} h'\right)$ that is equally as good as (h, c):*

$$MMU\left((h,c),R;w^{ref}\right)=\inf\left\{m|\,\exists\, h'\in[0,T]:(h,c)\,I\left(h',m+w^{ref}h'\right)\right\}. \tag{14}$$

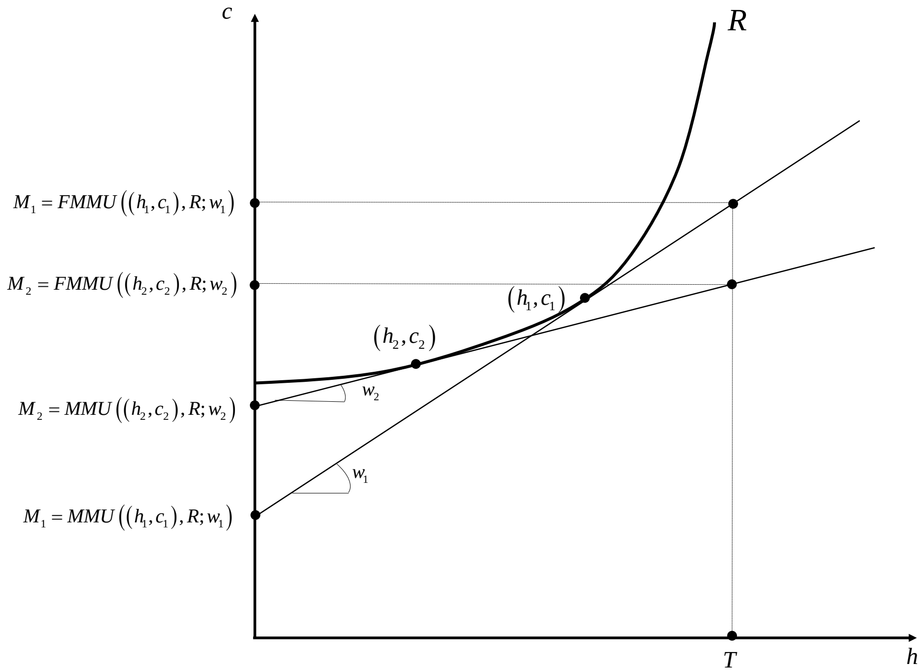
Definition 8. *Given an alternative (h,c), a reference wage w^{ref}, and a preference relation R, the full income money metric utility, say $FMMU\left((h,c),R;w^{ref}\right)$, is the minimal full income $Y := m+w^{ref}T$, such that there still exists a bundle $\left(h',Y-w^{ref}(T-h')\right)$ that is equally good as (h,c):*

$$FMMU\left((h,c),R;w^{ref}\right) = \inf\left\{Y|\,\exists h'\in[0,T]:(h,c)\,I\left(h',Y-w^{ref}(T-h')\right)\right\} \tag{15}$$

Note that

$$FMMU\left((h,c),R;w^{ref}\right)=MMU\left((h,c),R;w^{ref}\right)+w^{ref}T. \tag{16}$$

First of all, we observe that the unearned income version of the MMU ranks individuals differently from the full income version when individual specific wages are used. This is illustrated in Figure 3. This would not be the case if a common reference wage is used.



Note: the price of consumption c is normalised to one.

Figure 3: Ranking of FMMU does not equal ranking of MMU when wages differ.

In unitary models (e.g. Aaberge, Dagsvik, and Strøm, 1995, Aaberge, Colombino, Strøm, 2000, Creedy and Kalb, 2005 Section 2.2, and Dagsvik, Locatelli and Strøm 2009) as well as in collective models (Chiappori 2015, Chiappori and Meghir, 2016, Cherchye *et al.*, 2015, 2018) it is common practice to use market prices, *i.e.* wages earned in the actually chosen job, when applying this framework. As wages obviously differ across individuals, this is an instance of using different reference prices in money metric utilities.[9]

We illustrate now how the money metric utilities underlying the standard approach to calculating equivalent and compensating variations is subject to our critiques. Assume that preferences over (h, c) can be represented by a direct utility function $u(h, c)$. The true budget set is described by

$$c \leq f(w_g, h; I, \mathbf{s}), \tag{17}$$

where f embodies the tax–benefit system turning gross earnings $(w_g h)$, unearned income (I), into disposable income c, conditional on some other personal or household characteristics $\mathbf{s}$.

The compensating and equivalent variation will evaluate the welfare effect of changes in choices due to a reform in f on the basis of a conceptual experiment where one investigates the monetary compensation needed in order to restore original welfare or obtain the new level of welfare, in a situation where *ex post* and *ex ante*, workers would have been faced with a (virtual) linear budget constraint:

$$c \leq \omega h + \mu, \tag{18}$$

where ω is a virtual (reference) wage, and μ a virtual unearned income.

In case of continuous labour time, optimal labour time in this conceptual experiment is chosen from $[0, T]$. If labour time is considered as available only in discrete fractions, say $\{h_1, h_2, \ldots, h_k, \ldots, H_K\}$, optimal labour h in the conceptual experiment would be chosen from that set. The set of possible labour time regimes will in both cases be denoted by $\mathcal{H}$.

Let (h^0, c^0) and (h^1, c^1) denote the observed or simulated choice in baseline (0) with tax–benefit system f_0, and post reform (1) with tax–benefit system f_1. That is[10]:

$$\left(h^i, c^i\right) = \underset{h \in \mathcal{H}; c = f_i(w_g, h; I, \mathbf{z})}{\arg\max} u(h, c). \tag{19}$$

Furthermore, let $U^i = u(h^i, c^i)$.

Let the virtual wage ω^i $(i = 0, 1)$ be equal to $MRS(h^i, c^i) := \frac{\partial u(h^i, c^i)/\partial h}{\partial u(h^i, c^i)/\partial c}$. Then, the virtual unearned income equals $\mu^i := c^i - \omega^i h^i$. In such a case, both virtual wage and unearned income are endogenous (depend on an agent's choice) and depend on an individual's preferences.

Notice that $\max_{h \in \mathcal{H}} u(h, \omega^i h + \mu^i) \equiv u(h^i, c^i)$. As the solution to h is a function of ω^i and μ^i, we can write $\max_{h \in \mathcal{H}} u(h, \omega^i h + \mu^i)$ as $v(\omega^i, \mu^i)$, an indirect utility function. This yields:

$$v\left(\omega^i, \mu^i\right) = u\left(h^i, c^i\right). \tag{20}$$

[9] Another example is the use of Lindahl prices for public goods in the so–called sharing rule representation of the collective model (Cherchye *et al.*, 2011, 2015, 2018, Ferrando, 2018).

[10] We assume for simplicity of exposition that the maximisers in the next expression are unique.

Inverting around μ^i gives the expenditure function e:

$$\mu^i = e\left(\omega^i; u\left(h^i, c^i\right)\right). \tag{21}$$

So, μ^i is a money metric utility. The full income variant is defined as:

$$M^i := \mu^i + \omega^i T. \tag{22}$$

Let us now concentrate on the equivalent variation.[11] The equivalent variation EV is defined implicitly by:

$$\max_{h\in\mathcal{H}} u\left(h, \omega^0 h + \mu^0 - EV\right) = u\left(h^1, c^1\right). \tag{23}$$

As the solution h^* to the maximisation problem on the left hand side is a function of ω^0 and $\mu^0 - EV$, we can denote $\max_{h\in\mathcal{H}} u\left(h, \omega^0 h + \mu^0 - EV\right)$ by $v\left(\omega^0, \mu^0 - EV\right)$, resulting in an indirect utility function:

$$v\left(\omega^0, \mu^0 - EV\right) = u\left(h^1, c^1\right). \tag{24}$$

Inverting v around $\mu^0 - EV$ yields:

$$\mu^0 - EV = e\left(\omega^0; u\left(h^1, c^1\right)\right). \tag{25}$$

So, $\mu^0 - EV$ is a money metric utility of the post reform situation, using *ex ante* wages as the reference price for leisure (labour time). The full income variant is:

$$M\left(\omega^0, U^1\right) = \mu^0 - EV + \omega^0 T. \tag{26}$$

As *ex ante* choices differ, and people may have different preferences over (h, c), it turns out that the slope of the virtual budget constraint, that is the virtual wage (the marginal rate of substitution of c with respect to h in the chosen point), will not be identical across individuals. Therefore using μ^0 and $\mu^0 - EV$, the welfare measures underlying the EV, is subject to the criticisms raised in Section 3. In our empirical application (Section 6), we illustrate that these problems may be all but negligible in a real world situation.

5 Implications for measuring welfare differences

In much applied work, one investigates the characteristics of the biggest winners or losers of a reform. This requires interpersonal comparisons of welfare differences. We investigate in this section some consequences of the Same Preference Principle and the Nested Contour Criterion for measuring and

[11] Similar arguments can be made for the compensating variation.

comparing welfare differences, and show that the equivalent variation does not satisfy these implied properties. Again, the argument equally well applies to the compensating variation. We continue to work within the labour consumption framework, but all our statements apply to the more general framework of Section 2.

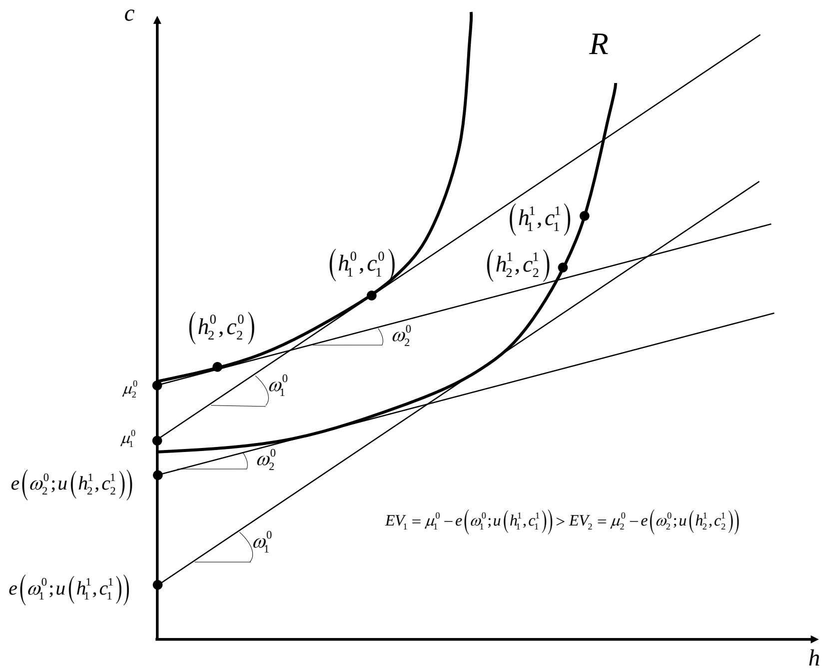
From Equations (21) and (25) we can derive the usual definition of the equivalent variation, which is a measure of welfare differences:

$$EV = \mu^0 - e\left(\omega^0; u\left(h^1, c^1\right)\right) = e\left(\omega^0; u\left(h^0, c^0\right)\right) - e\left(\omega^0; u\left(h^1, c^1\right)\right). \tag{27}$$

SPP has the following implications for measuring welfare differences between two situations, say 0 and 1. Let there be two persons, 1 and 2, with identical preferences, say R. Let it be the case that $(h_1^0, c_1^0)\ I\ (h_2^0, c_2^0)$ and $(h_1^1, c_1^1)\ I\ (h_2^1, c_2^1)$, as for example in Figure 4. Then it follows from SPP that the following equality should hold:

$$W_1\left(\left(h_1^0, c_1^0\right), R\right) - W_1\left(\left(h_1^1, c_1^1\right), R\right) = W_2\left(\left(h_2^0, c_2^0\right), R\right) - W_2\left(\left(h_2^1, c_2^1\right), R\right). \tag{28}$$

Figure 4 shows that the equivalent variation as defined in Equation (27) violates this equality.



Note: the price of consumption c is normalised to one.

Figure 4: EV does not satisfy SPP.

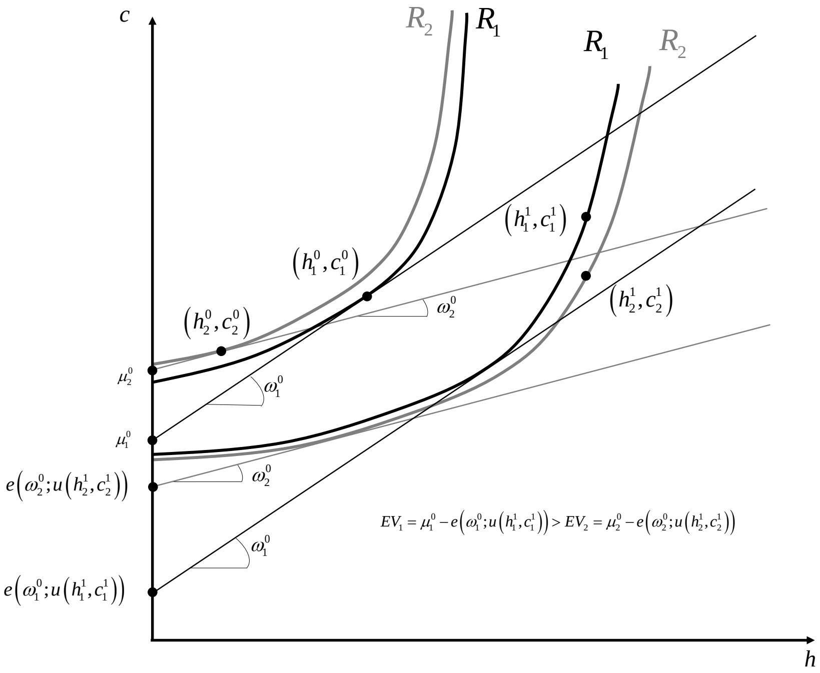
Next, turn to the implications of NCC for measuring welfare differences. Suppose $LC\left(\left(h_1^0, c_1^0\right), R_1\right) \cup UC\left(\left(h_2^0, c_2^0\right), R_2\right) = \varnothing$ and $LC\left(\left(h_2^1, c_2^1\right), R_2\right) \cup UC\left(\left(h_1^1, c_1^1\right), R_1\right) = \varnothing$, as *e.g.* in Figure 5. Then, NCC implies that

$$W_2\left(\left(h_2^0, c_2^0\right), R_2\right) > W_1\left(\left(h_1^0, c_1^0\right), R_1\right) \text{ and } W_1\left(\left(h_1^1, c_1^1\right), R_1\right) > W_2\left(\left(h_2^1, c_2^1\right), R_2\right), \tag{29}$$

and therefore it should hold that

$$W_2\left(\left(h_2^0, c_2^0\right), R_2\right) - W_2\left(\left(h_2^1, c_2^1\right), R_2\right) > W_1\left(\left(h_1^0, c_1^0\right), R_1\right) - W_1\left(\left(h_1^1, c_1^1\right), R_1\right). \tag{30}$$

Figure 5 shows that EV violates this inequality.



Note: the price of consumption c is normalised to one.

Figure 5: EV does not satisfy NCC.

Similar critiques hold for the compensating variation. If one wants to accept SPP or NCC as sound principles of interpersonal welfare comparisons, the equivalent and compensating variation are not good measures for the purpose of making interpersonal comparisons of welfare differences, *e.g.* for comparing the extent to which individuals are affected by policy reforms.

6 Empirical illustration

In order to illustrate the relevance of results 1–3, we use a dataset in the labour consumption context. For the individuals in the dataset, we calculate money metric utilities using estimated preferences and next we show that the axioms SPP, NCC and NOSP are violated by a considerable proportion of pairs of individuals. As NOSP is a strengthening of NCC, we focus on SPP and NCC. In what follows we will use the money metric utility in the labour consumption context as in Definition 7 and full income money metric utility as in Definition 8. We will refer to the (full income) MMU where reference wages are equal to own wages by "own wage MMU" resp. "own wage FMMU".

For the SPP, we will count the number of pairs of individuals with identical estimated preferences, where the bundle of the first individual is preferred over the bundle of the second individual by their common preference relation, and despite this, the own wage money metric utility is higher for the second individual than for the first.

For the NCC, we will analogously count the number of pairs of individuals, where the upper contour set of the first individual (with her observed bundle) and the lower contour set of the second individual (with her observed bundle) are disjoint, and despite this, the own wage money metric utility is higher for the second individual than for the first.

We thus show that the violations of SPP and NCC are not merely theoretically possible issues, but can be empirically relevant as well.

Finally we will compare the rankings of the individuals with respect to the two versions of the own wage money metric utilities and two money metric utilities with (two different) common reference wages.

6.1 Data and estimated preferences

The data for this empirical illustration stem from the Belgian version of the SILC (Statistics on Income and Living Conditions) 2015 survey. It is a dataset at the micro (*i.e.* individual) level, which includes, among others, information on personal characteristics, detailed information on income, labour time, poverty and other living conditions.

From this dataset, a sub–sample was taken that only contains persons who are the sole adult available for the labour market in their household and who are between 18 and 64 years old. Self–employed individuals and employers are excluded. The final sub–sample contains 624 females and 526 men. Summary statistics of the data can be found in Table 1.

Preferences were estimated by Maximum Likelihood for a Random Utility Random Opportunity model as in Capéau et al. (2018). Preferences are all of the Box–Cox type, but they are allowed to be gender specific, and marginal rates of substitution may depend on age, education, region and the number of children (these variables are denoted by the vector $\mathbf{r}$):

$$U(h,c) = \beta_c \frac{c^{\alpha_c} - 1}{\alpha_c} + \boldsymbol{\beta}_l' \mathbf{r} \frac{(\frac{T-h}{T})^{\alpha_l} - 1}{\alpha_l}, \tag{31}$$

where $\frac{T-h}{T}$ represents the normalised amount of leisure time.

The estimated coefficients can be found in Table 2. The exponents, α_c and α_l, determine the curvature of the indifference curves. The lower these coefficients, the more the shape of the indifference curves bends towards a rectangular form. For given α_c and α_l, β_c and $\boldsymbol{\beta}_l'\mathbf{r}$ tilt the indifference curves (they become steeper with lower β_c and higher value of $\boldsymbol{\beta}_l'\mathbf{r}$). Therefore, preferences with identical values

Table 1: Descriptive statistics SILC sub–sample.

Description	Female	Male
Age (years)	43.8	42.2
Dependent children (%)		
0 – 3 years	6.0	0.8
4 – 6 years	6.8	1.7
7 – 9 years	8.5	1.1
Experience (years)	18.0	19.1
Education (%)		
Low	21.1	19.4
Middle	35.2	37.8
High	43.6	42.8
Residence (%)		
Brussels	28.6	24.0
Flanders	39.4	44.1
Wallonia	32.0	31.9
Participation rate (%)	70.5	75.1
Hours worked (hours/week)		
Unconditional	24.8	30.4
Conditional on working	35.2	40.4
Hourly wage (euro)	21.0	22.2
Disposable income (euro/month)	2083.3	2153.73
Number of observations	644	526

Calculations based on a sub–sample of the Belgian version of SILC 2015. Disposable incomes are calculated by means of EUROMOD version G4.0+.

for α_c and α_l satisfy the single crossing condition. Thus, in our application males' preferences are single crossing, and the same holds true for the females. Multiple crossings occur when comparing a male with a female's preference map.

Table 2: Estimates of preference parameters within a RURO model.

Description	Female	Male
Box-Cox utility function		
constant consumption	5.48	5.21
exponent consumption	0.34	0.43
exponent leisure	-5.68	-3.82
Leisure covariates		
constant	84.38	1.50
log(age)	-44.47	1.04
$\log(\text{age})^2$	6.12	-0.04
child 0 – 3	-0.53	·
child 4 – 6	0.16	-1.58
child 7 – 9	-0.08	-1.14
Brussels	0.39	0.66
Wallonia	-0.22	0.08
low education	0.60	-0.13
high education	-0.85	-0.76

Estimates are based on a sub–sample of the Belgian version of SILC 2015.
More details can be found in Capéau *et al.* (2018).

For observed values (h^0, c^0), we calculated MMU's for a common reference wage w^r, say $MMU(h^0, c^0; w^r)$ by numerically solving the following problem:

$$\max_{h\in[0,T]} u(h, m + w^r h) = u\left(h^0, c^0\right). \tag{32}$$

The value of m solving this equation is $MMU(h^0, c^0; w^r)$.

For the individual specific reference price ω^i, it holds that:

$$MMU\left(h^0, c^0; \omega_i\right) = c^0 - \omega^i h^0, \tag{33}$$

where ω^i is (minus) the marginal rate of substitution of consumption for leisure,

$$\omega^i = \frac{\boldsymbol{\beta}_l' \mathbf{r} \left(T - h^0\right)^{\alpha_l - 1} / T^{\alpha_l}}{\beta_c \left(c^0\right)^{\alpha_c - 1}}. \tag{34}$$

To obtain the full income variants of $MMU\left(h^0, c^0; w^r\right)$ and $MMU\left(h^0, c^0; \omega_i\right)$, one has to add $w^r T$, respectively $\omega^i T$, to these values.[12]

We do not take into account the discrete choice aspect of our positive model as we consider labour as a continuous variable in order to carry out the conceptual experiment that puts monetary value on welfare actually obtained. Even so, actual welfare obtained will be affected by the real choice environment which possibly embodies demand side restrictions and discreteness of available choice options, and will as such have an impact on our welfare calculations. We also neglected random preference variation embodied in the random terms of utilities in a discrete choice model, which we consider to be a real shortcoming.[13]

6.2 Conflict 1: Same Preference Principle (SPP)

In this subsection we study the violations of SPP. Therefore, we divide the data into groups of individuals with identical preferences. The distribution of the sizes of the so constructed groups is represented in the second column of Table 3.

Table 3: Total number of conflicts by group size of individuals with identical preferences.

size of the group	number of groups of this size	total number of comparisons	total number of conflicts MMU	total number of conflicts FMMU
1	355	0		
2	154	154	33	40
3	85	255	61	54
4	30	180	48	50
5	14	140	34	30
6	6	90	24	23
8	2	56	21	9
10	1	45	24	11
total		920	245	217

For the 355 persons with idiosyncratic preference, SPP has no bite. For the other groups, we pairwise compare individuals and count the number of comparisons and conflicts for each group. A conflict is defined here as a pair of individuals where the bundle of the first individual is preferred over the bundle of the second individual by their common preference relation, and nevertheless, the own wage money metric utility is higher for the second individual than for the first. A summary of the

[12] For numerical reasons we restricted the consumption set X to be equal to $[0, 112\text{x}4.33] \times \mathbb{R}_+$ in the current evaluation exercise. This means that we did not add (168x4.33)w^r or (168x4.33)ω_i to the MMU, but (112x4.33)w^r or (112x4.33)ω_i.

[13] For the implications of true discreteness in choice and random preference heterogeneity on measuring welfare, see *e.g.* the work of Small and Rosen (1981), McFadden (1999), Dagsvik and Karlström (2005), Creedy and Kalb (2005, Section 3), Creedy, Hérault and Kalb (2011), and Bhattacharya (2015).

results can be found in Table 3.

When calculating the total number of comparisons made between pairs of individuals with identical preferences and conflicts in the above defined sense, we get a conflict proportion for the own wage MMU of more than 26% and a conflict proportion for the own wage FMMU of more than 23% as can be seen in Table 3. It turns out that the failure to satisfy the Same Preference Principle is not only a theoretical curiosity, but it is empirically relevant as well.

6.3 Conflict 2: Nested Contour Criterion (NCC)

Besides the Same Preference Principle (SPP), the Nested Contour Criterion (NCC) is also not satisfied by the own wage versions of the MMU, as was shown in Result 2. In order to assess the empirical relevance of this second finding, we again pairwise compare all individuals in the dataset and count the conflicts. A conflict is defined here as a pair of individuals, where the upper contour set of the first individual (with her observed bundle) and the lower contour set of the second individual (with her observed bundle) are disjoint, and nevertheless, the own wage money metric utility is higher for the second individual than for the first.[14]

Table 4: NCC: Overall number of comparisons, conflicts and proportion of conflicts.

total number of comparisons	683 865
number of dominances of indifference curves	316 429
number of conflicts MMU	76 514
number of conflicts FMMU	83 146
proportion conflicts MMU compared to dominances (%)	24,2
proportion conflicts FMMU compared to dominances (%)	26,3

Table 4 summarises the results. We give the absolute numbers of total comparisons and total conflicts, and also the proportion of conflicts compared to total number of pairs where one individual's indifference curve dominates the other. One can see that also for the NCC, in more than 24% of the comparisons where there is dominance, a conflict arises for MMU, and in more than 26% of the comparisons where there is dominance, a conflict arises for FMMU. Again, the violation of NCC turns out not to be limited to some concocted examples which are empirically irrelevant.

6.4 Rank reversals

Given that the normative principles underlying money metrics with common reference prices are different from those using individually specific prices, we now investigate its implications for the

[14] We determined the disjointness of upper and lower contour sets by numerically investigating whether the indifference curve in (h, c)–space of one individual lies above that of another one. In line with the remark in note 12, we limited our investigation to the hours range of $[0, 112]$.

persons considered to be relatively well and worse off by both type of metrics.

Thereto we will use rank plots. In these plots, we do not plot the value of the MMU or FMMU itself, but the relative position or rank a person obtains according to one measure *vis–à–vis* the rank obtained according to another measure.

In Figure 6 we shed light on the impact of the choice of the reference wage on the ranking. The horizontal axis contains the rank numbers for the MMU based on a reference wage equal to the first quartile ($Q1$) in the observed wage distribution. Along the vertical axis the corresponding rank of that person according to the MMU based on a reference wage equal to the third quartile ($Q3$) in the wage distribution is plotted. Dots are shaded according to the position of a person in the wage distribution (darker indicates a lower wage position, lighter a higher wage position). Figure 6 shows that both rankings, though not identical, are considerably consistent. The correlation between the ranks is equal to 0.98. Moreover there is no clear pattern in the wage of those being ranked high or low according to both measures. Note that exactly the same picture would have been obtained when using the FMMU instead of the MMU, since in Section 4 we showed that both alternatives result in the same ranking when using common reference wages.

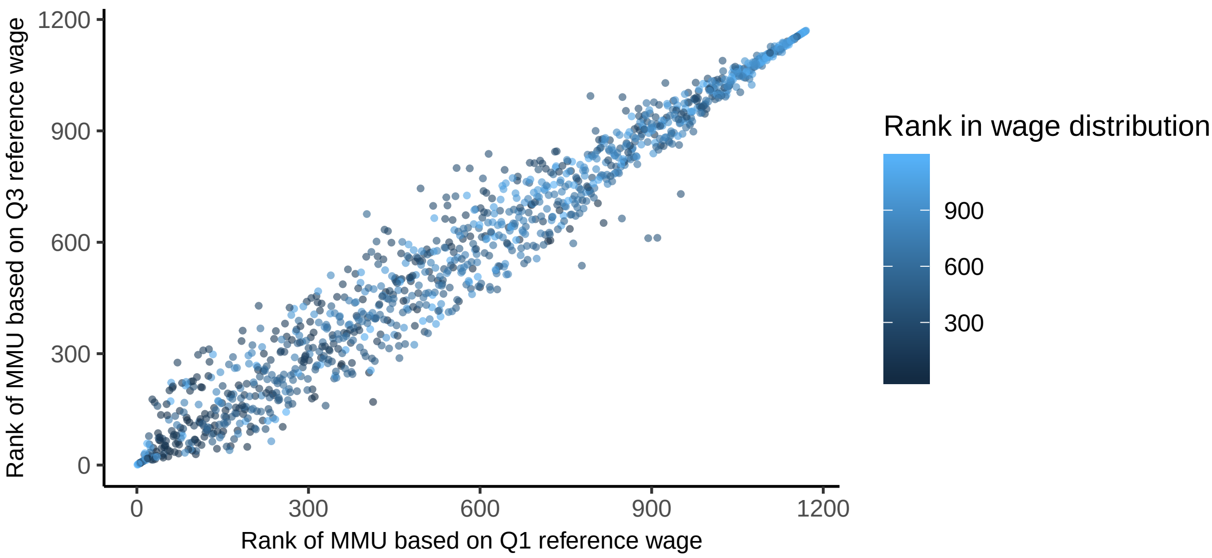


Figure 6: Plot of ranks according to MMU with $w^r = Q3$ against MMU with $w^r = Q1$.

Next, in Figures 7 and 8, we compare the ranking of own wage MMU with the ranking of MMU based on the two common values for the reference wages, $Q1$ and $Q3$. The consistency between the own wage MMU and the two reference wage MMUs turns out to be considerably smaller than in Figure 6. The correlation between the own wage MMU and the MMU's with common reference wages drops to 0.63. There is a tendency for persons with high wages to be ranked at the bottom when using an own wage MMU while they were at a much higher place in the well–being distribution when using common reference wages.

We analogously compare the own wage FMMU with the (F)MMU with reference wages in Figures 9

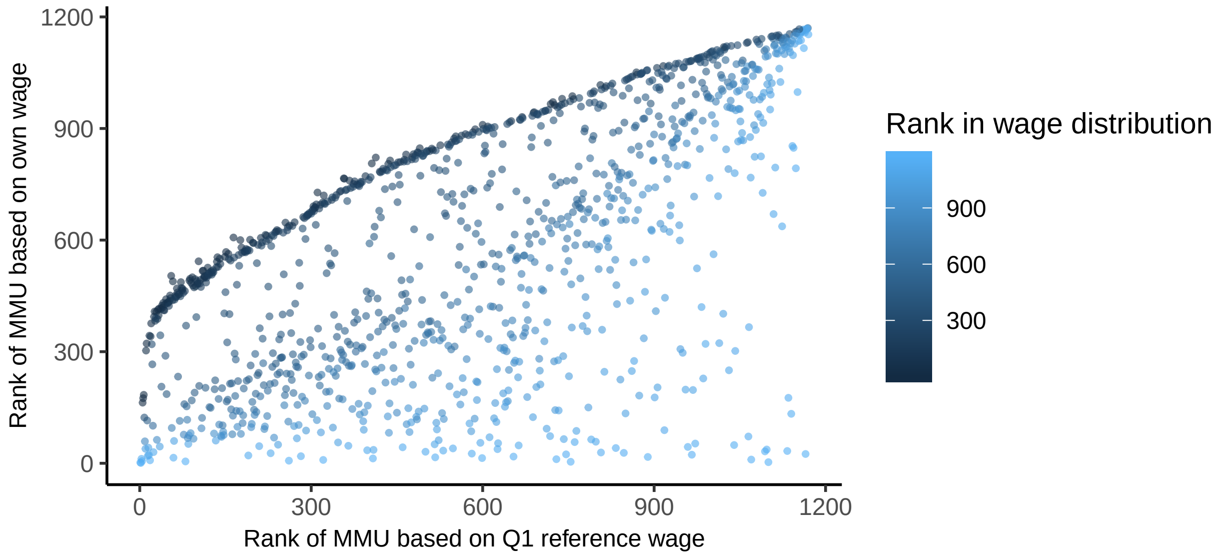


Figure 7: Plot of ranks of own wage MMU against MMU with $w^r = Q1$.

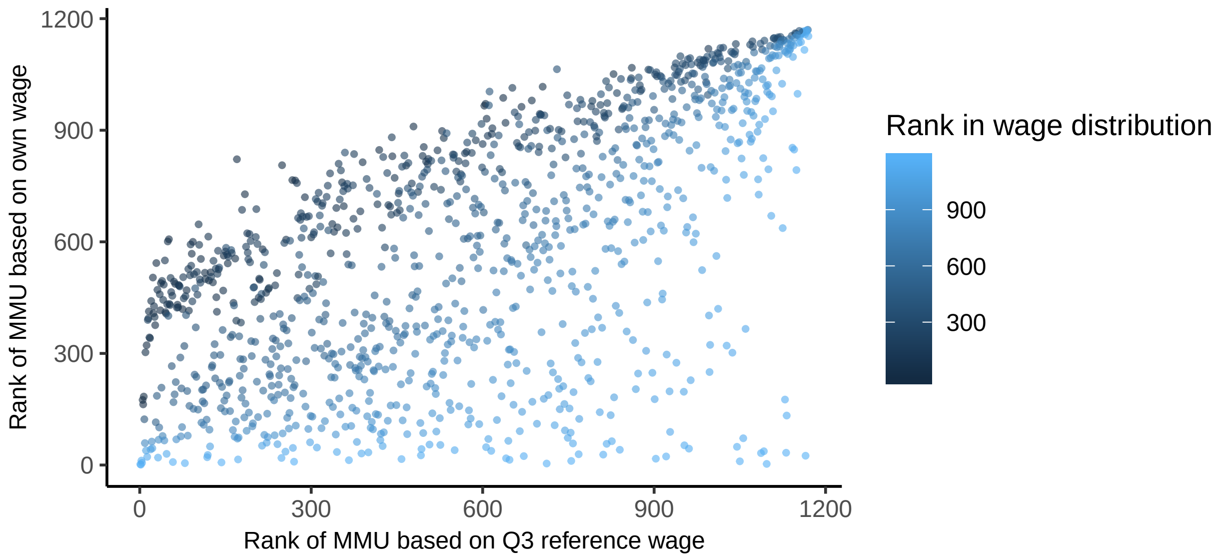


Figure 8: Plot of ranks of own wage MMU against MMU based on $w^r = Q3$.

and 10. In both graphs, we maintain the rank according to the common reference wage MMU on the horizontal axis, as this rank is the same as the ranks of the corresponding FMMU. Also these figures illustrate that the consistency between the own wage FMMU and the two reference wage (F)MMU's is considerably smaller than in Figure 6. Correlations between the own wage FMMU and the (F)MMU's with common reference wages again drop to 0.63–0.64. Now there is a tendency for persons ranked at the top of the wage distribution, to be ranked lowly according to the FMMU with common reference wage, and highly according to the own wage FMMU. This stands to reason, as the value of the time endowment, $\omega^i T$, added to the MMU in case of the own wage FMMU is higher for persons with high wages, while it is a constant for the common reference wage FMMU.

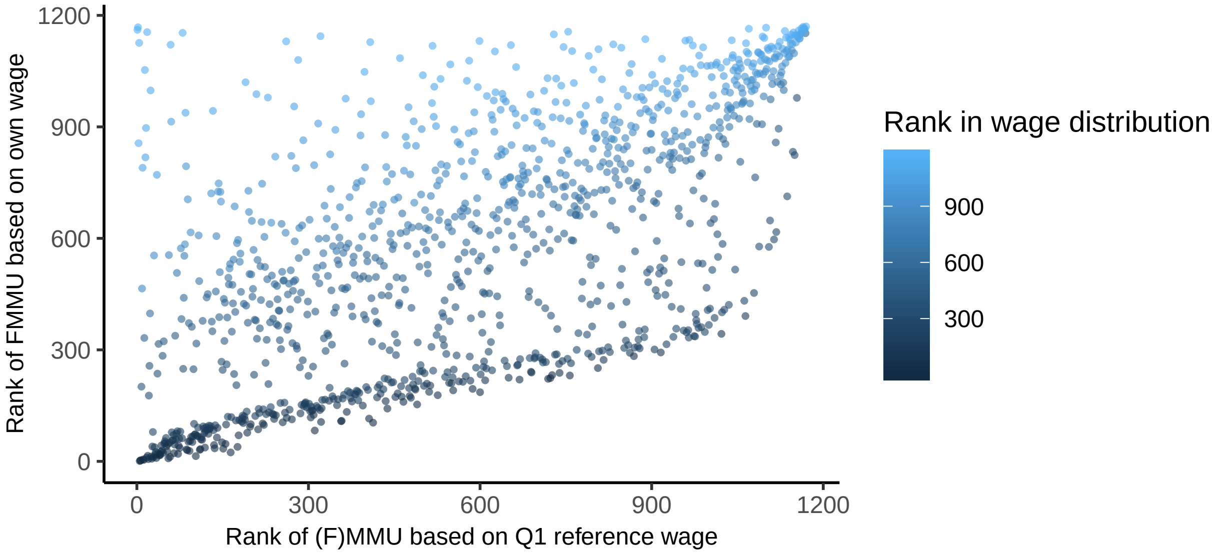


Figure 9: Plot of ranks of own wage FMMU against (F)MMUs with $w^r = Q1$.

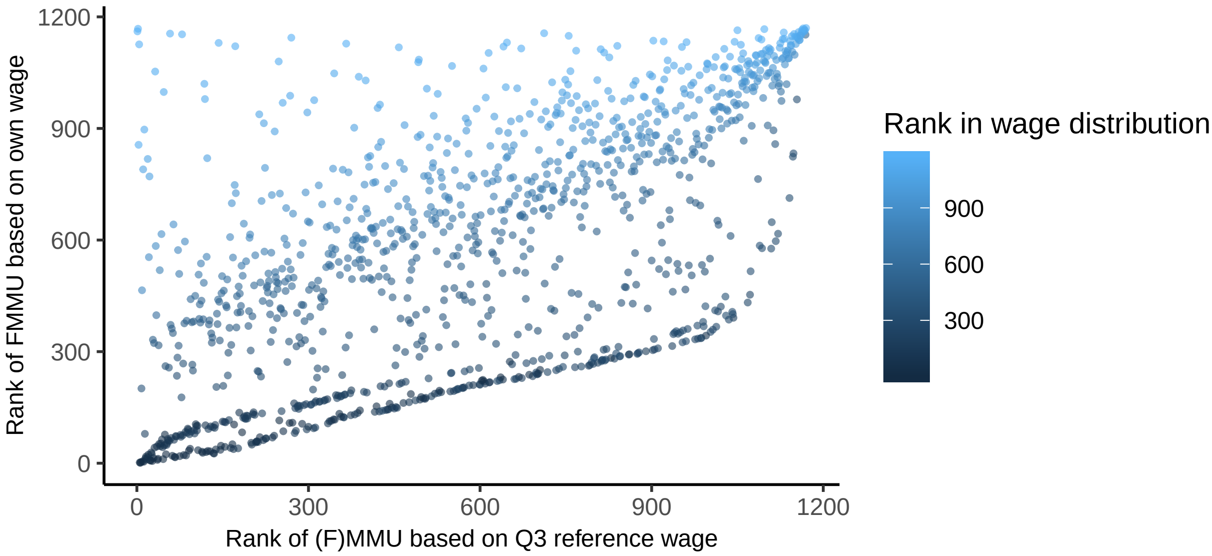


Figure 10: Plot of ranks of own wage FMMU against (F)MMUs with $w^r = Q3$.

6.5 Properties of own wage welfare rankings

As the own wage MMU and FMMU do not yield the same ranking of individuals, we compare their rankings in Figure 11. It turns out that there is a tendency for the high wage persons to be ranked higher by the FMMU while low wage individuals are more often ranked higher by the MMU: a considerable number of light shaded dots (high wages) appear in the upper right corner (ranked high by FMMU and low by MMU), while the lower right corner is predominantly occupied by darkly shaded dots (low wages). Again this is due to the fact that the FMMU adds the value of time endowment, $\omega^i T$, to the MMU, which is higher for persons with high wages. On top of that, the MMU of a low wage person with the same labour time consumption bundle as a high wage person is by definition higher. We conclude from this analysis that the actual wage earned has a considerable influence on the ranking according to (F)MMU's. We feel this is unattractive, in the sense that the welfare effect of high wages is already fully reflected in the better opportunities it offers in terms of leisure and consumption, and should not be exacerbated by valuing these opportunities at a higher price. If two persons obtain the same leisure consumption bundle, the fact that they have different wages is no reason *per se* to rank them differently. Within our model, only differences in their intensity of preferences for leisure can (or cannot, depending on your normative position) give rise to different rankings of such persons.

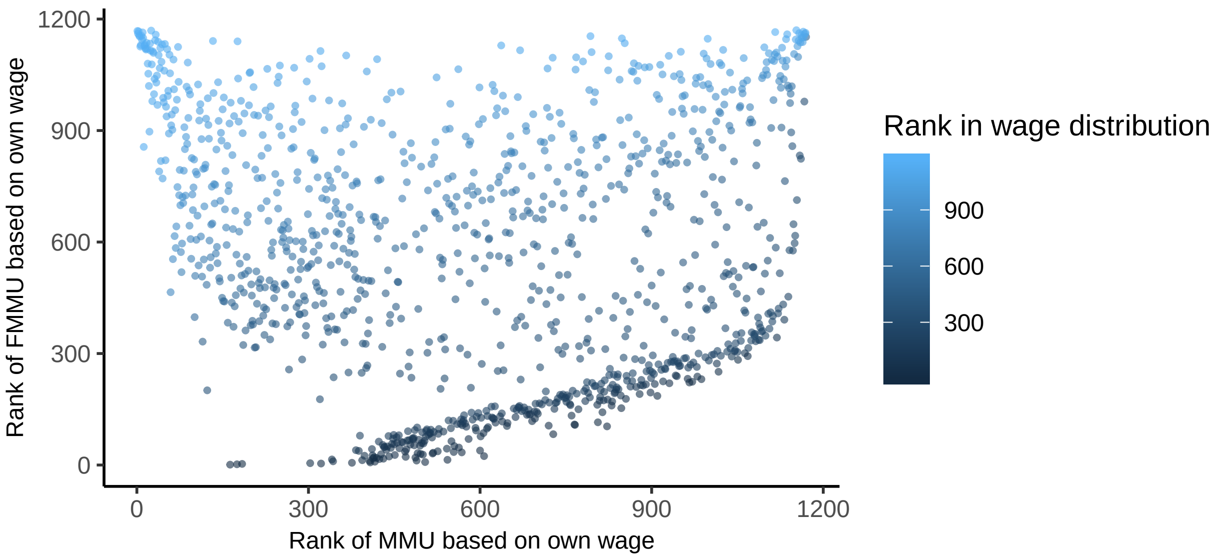


Figure 11: Plot of ranks of own wage MMU against own wage FMMU.

In Figures 12 and 13 we extend our investigation of the properties of the differences in ranking according the own wage MMU and FMMU by including information on labour time in the picture. Thereto we compare the wage ranking (horizontal axis) with the ranking according the own wage (F)MMU (the dots). The shading now reflects the ranking in the labour time distribution: lighter dots are those of persons working longer. Figure 12 confirms that there is a weak negative relation between the wage and MMU's, which is also reflected in the correlation coefficient of -0.33.

The shading pattern suggests moreover that there is a tendency for persons who work longer to be ranked lower according to the own wage MMU. The correlation coefficient between both rankings is -0.45.

Figure 12: Plots of wage ranking against own wage MMU ranking.

In turn, Figure 13 confirms that there is a systematic positive relation between wage ranks and ranks according to the own wage FMMU. The correlation coefficient is 0.95. Moreover, the shading of the dots reveals that there is a positive relation between the observed hours worked and the own wage FMMU ranking. The correlation coefficient is 0.82. Again, such high correlations in absence of a precise indication of the extent to which such differences in labour time choices are due to differences in terms of intensity of preference for leisure or rather due to actual wage differences, are not very attractive. By fixing the reference wage one disentangles both aspects, and safeguards only the preference differences factor. Admittedly, information on preference differences is limited (the indifference curves are only compared at a particular point), and may be difficult to obtain in practice.

7 Conclusion

We have shown that the common practice of using different reference prices for different persons when using the money metric utilities as a measure of individual well–being, leads to the violation of several normative principles of individual well–being measurement: the Same Preference Principle, the Nested Contour Criterion, and the Nested Opportunity Set Property. The equivalent and compensating variation violate some immediate consequences of the Same Preference Principle and the Nested Contour Criterion for interpersonal comparisons of welfare differences. Therefore, such a practice is in need of another normative foundation.

Figure 13: Plot of wage ranking against own wage FMMU ranking.

Moreover, we show that the violation of these principles is not merely a theoretical issue, but can be empirically relevant as well.

None of these problems occur when switching to a common reference price for all. Of course, the next problem is then to choose the appropriate reference price. We think this choice too should be based on normative grounds (see *e.g.* Capéau *et al.*, 2016 for an attempt in this direction). Our empirical results are comforting in the sense that in the case at hand, the choice of the reference wage seemed to have only a weak impact on the welfare ranking.

To focus on the role of a common versus individual specific reference price vector, we neglected the discrete choice character of our empirical labour supply model, and the random preference differences. Concerning the first, we believe that the welfare effect of restricted choice sets is embodied in the actual choice people make. It will yield lower welfare in presence than in absence of such restrictions. We have the feeling that labour time can be conceived of as a continuous variable in the evaluation of counterfactuals necessary for interpersonally comparable welfare measurement. This might however not be the case for other discrete choice contexts, such as the choice between transport modes. Concerning the second, we refer to the work of McFadden (1999), Dagsvik and Karlström (2005), Creedy, Hérault, and Kalb (2011), and Carpantier and Sapata (2016) for some reflections on paths to integrate random preference differences in money metrics based welfare measures.

Finally, our paper contains an opportunity based motivation of MMU's with common reference prices by means of the Nested Opportunity Set Property. We pinpointed a potential conflict between this property and a more preference dependent way of welfare measurement embodied in a stronger version of the Nested Contour Criterion that encompasses the Same Preference Principle. This potential trade off between strengthening the Nested Contour Criterion and the opportunity motivation of well–being measurement requires further investigation.

References

[1] Aaberge R., J.K. Dagsvik, & S. Strøm (1995) "Labor supply responses and welfare effects of tax reforms," *Scandinavian Journal of Economics*, Vol. 97(4), 635–659.

[2] Aaberge R., U. Colombino, & S. Strøm (2000) "Labor supply responses and welfare effects from replacing current tax rules by a flat tax: Empirical evidence from Italy, Norway and Sweden," *Journal of Population Economics*, Vol. 13(4), 595–621.

[3] Bhattacharya D. (2015) "Nonparametric welfare analysis for discrete choice," *Econometrica*, Vol. 83(2), 617–649.

[4] Blackorby C. & D. Donaldson (1990), "A review article: The case against the use of the sum of compensating variations in cost–benefit analysis," in *Canadian Journal of Economics – Revue canadienne d'Économique*, Vol. 13(3), 471–494.

[5] Blackorby C., F. Laisney, & R. Schmachtenberg (1994), "Ethically–consistent welfare prescription are reference price–independent," in: R. Blundell, I. Preston, and I. Walker (eds.) *The Measurement of Household Welfare*, Cambridge University Press: Cambridge, 247–264.

[6] Blundell R., I. Preston, & I. Walker (1994), "An introduction to applied welfare analysis," in: R. Blundell, I. Preston, and I. Walker (eds.) *The Measurement of Household Welfare*, Cambridge University Press: Cambridge, 1–50.

[7] Boadway R.W. (1974) "The welfare foundations of cost–benefit analysis," *Economic Journal*, Vol. 84 No.334, 926–939.

[8] Bossert W., P.K. Pattanaik, & Y. Xu (1994, "Ranking opportunity sets: An axiomatic approach," *Journal of Economic Theory*, Vol. 63(2), 326–345.

[9] Capéau B., A. Decoster, & D. Neumann (2016) "Some properties of the choice of reference wage for the equivalent income metric," *mimeo.*

[10] Capéau B., A. Decoster, S. Maes, & T. Vanheukelom (2018) "Piecemeal modelling of the effects of joint direct and indirect tax reforms," *EUROMOD Working Papers* EM14/18.

[11] Carpantier J.F. & C. Sapata (2016), "Empirical welfare analysis: when preferences matter," *Social Choice and Welfare*, Vol. 46(3), 521–542.

[12] Cherchye L., S. Cosaert, B. De Rock, P. J. Kerstens, & F. Vermeulen (2018), "Individual welfare analysis for collective households," *Journal of Public Economics*, Vol. 166, 98–114.

[13] Cherchye L., B. De Rock, A. Lewbel, & F. Vermeulen (2015), "Sharing rule identification for general collective consumption models," *Econometrica*, Vol. 83(5), 2001–2041.

[14] Cherchye L., B. De Rock, & F. Vermeulen (2011), "The revealed preference approach to collective consumption behaviour: Testing and sharing rule recovery," *Review of Economic Studies*, Vol. 78(1), 176–198.

[15] Chiappori P.–A. (2015) "Welfare and the household," in: M.D. Adler & M. Fleurbaey (eds.), *The Oxford Handbook of Well–being and Public Policy*, Oxford University Press: New York, 785–843.

[16] Chiappori P.–A. & C. Meghir (2016) "Intrahousehold Inequality," in: A.B. Atkinson & F. Bourguignon (eds.), *Handbook of Income Distribution*, Vol. 2B, North-Holland: Amsterdam, 1369–1418.

[17] Creedy J., N. Hérault, & G. Kalb (2011) "Measuring welfare changes in behavioural micro simulation modeling: Accounting for the random utility component," *Journal of Applied Economics*, Vol. 14(1), 5–34.

[18] Creedy J. & G. Kalb (2005) "Measuring welfare changes in labour supply models," *The Manchester School*, Vol. 73(6), 664—685.

[19] Donaldson D. (1992) "On the aggregation of money measures of well–being in applied welfare analysis," *Journal of Agricultural and Resource Economics*, Vol. 17(1), 88–102.

[20] Dagsvik J.K. & A. Karlström (2005), "Compensating variation and Hicksian choice probabilities in random utility models that are nonlinear in income,", *The Review of Economic Studies*, Vol.72(1), 57–76.

[21] Dagsvik J.K., M. Locatelli, & S. Strøm (2009) "Tax reform, sector–specific labor supply and welfare effects," *Scandinavian Journal of Economics*, Vol. 111(2), 299–321.

[22] Decancq K., M. Fleurbaey, & E. Schokkaert (2015), "Happiness, equivalent incomes and respect for individual preferences," *Economica*, Vol. 82(1), 1082–1106.

[23] Decancq K., M. Fleurbaey, & E. Schokkaert (2017), 'Well–being inequality and preference heterogeneity," *Economica*, Vol. 84(334), 210–238.

[24] Decancq K., M. Fleurbaey, & F. Maniquet (2015), "Multidimensional poverty measurement with individual preferences," CORE DP 2015/8 (forthcoming in Journal of Economic Inequality).

[25] De Sadeleer L. (2017), "A study of well–being and inequality with preference based individual welfare metrics applied to the income–leisure choice for singles and couples," Master thesis, Department of Economics KU Leuven.

[26] Ferrando M. (2018) "Intra–household allocations with public goods: A new source of gender inequality," *mimeo.*

[27] Fleurbaey M. (2008), *Fairness, Responsibility, and Welfare*, Oxford University Press: Oxford.

[28] Fleurbaey M. (2011), “Willingness–to–pay and the equivalence approach,” *Revue d’Économie Politique*, Vol. 121(1), p. 35–58.

[29] Fleurbaey M. & D. Blanchet (2013), *Beyond GDP: Measuring Welfare and assessing Sustainability*, Oxford University Press: Oxford.

[30] Fleurbaey M. & F. Maniquet (2011), *A Theory of Fairness and Social Welfare*, Cambridge University Press: Cambridge.

[31] Fleurbaey M. & F. Maniquet (2017), “Fairness and well–being measurement,” *Mathematical Social Sciences*, Vol.90, 119–126.

[32] Fleurbaey M. & F. Maniquet (2018), “Inequality–averse well–being measurement,” *International Journal of Economic Theory*, Vol.14, 35–50.

[33] Gul F. & W. Pesendorfer (2001), “Temptation and self–control,” *Econometrica*, Vol. 69(6), 1403–1435.

[34] King M.A. (1983), “Welfare analysis of tax reforms using household data,” *Journal of Public Economics*, Vol. 21(2), 183–214.

[35] King M.A. (1987), “The empirical analysis analysis of tax reforms,” in T.F. Bewley (ed.) *Advances in Econometrics: Fifth World Congress*, Vol. 2, Cambridge University Press: Cambridge, 61–90.

[36] McFadden D. (1999) “Computing willingness–to–pay in random utility models,” in: J.R. Melvin, J.C. Moore, & R. Riezman (eds.) *Trade, Theory and Econometrics. Essays in honor of John S. Chipman*, Routledge: London and New York, 253–274.

[37] Pauwels W. (1978), “The possible perverse behavior of the compensating variation as a welfare ranking,” Zeitschrift für Nationalökonomie/Journal of Economics, Vol. 38, Issue 3–4, 369–378.

[38] Pauwels W. (1986), “Correct and incorrect measures of the deadweight loss of taxation,” Public Finance – Finances Publiques, , vol. 41(2), 267–276.

[39] Roemer J.E. & A. Trannoy (2015), “Equality of opportunity,” in: A.B. Atkinson & F. Bourguignon (eds.), *Handbook of Income Distribution*, Vol. 2A, North-Holland: Amsterdam, 217–300.

[40] Roemer J.E. & A. Trannoy (2016), “Equality of opportunity: Theory and measurement,” *Journal of Economic Literature*, Vol. 54(4), 1288–1332.

[41] Small K.A. & H.S. Rosen (1981) “Applied welfare economics with discrete choice models,” *Econometrica* Vol. 49(1), 105–130.

[42] Weymark J.A. (1985), “Money metric utility functions,” *International Economic Review*, Vol. 26(1), 219–232.