Welfare and Inequality Comparisons for Uni- and Multi-dimensional Distributions of Ordinal Data

Frank A. Cowell, Martyna Kobus, Radosław Kurek

STICERD
London School of Economics
Houghton Street, London WC2A 2AE
f.cowell@lse.ac.uk

Institute of Economics
Polish Academy of Sciences
Nowy Swiat 72, 00-330 Warsaw
mkobus@inepan.waw.pl

Vistula University
Stokolosy Street 2
02-787 Warsaw
rkurek@vistula.edu.pl

January 2017

Abstract

Decision makers and social planners are often faced with a problem of evaluating distributions of ordinal variables i.e. variables for which there are no numbers but only the ordering, such as, for example, self-reported health status, life satisfaction, working environment, quality of public goods, living conditions etc. Standard tools, namely, stochastic dominance, and inequality and risk measures, produce conclusions that can be reversed depending on the cardinalisation of an ordinal indicator which is arbitrary. Utilising the notion of integration on partially ordered sets we extend the well-known Hardy et al. (1934) result to an ordinal setting, both univariate and multivariate.

Acknowledgements: This work was supported by National Science Centre in Poland, grant 2013/09/D/HS4/01029.

Keywords: Ordinal data; Inequality and Risk Orderings; Welfare orderings; Association

JEL codes: D30; D63; D81; I31

1 Introduction

The question of whether one distribution is better than another in some normative sense has been for long addressed in economics, drawing on a parallel literature in mathematics. Typically such comparisons are made based on a result that states "distribution f is better than distribution g if f can be obtained from g by means of a finite sequence of elementary transfers," or "...if f stochastically dominates g" or "...if all utilitarian decision makers who exhibit certain regularity in their preferences would choose f over g." How one defines elementary transfers, dominance conditions or regularity properties of preferences determines whether we compare distributions according to welfare, inequality or both. For inequality and risk[1] the dominance criterion is second order stochastic dominance; regularity of utilitarian decision makers' preferences means that they all have non-decreasing and concave utility functions.[2] Similar results have been established for the case when f, g are multidimensional.[3] However, these classic results assume that the attribute that is being considered is cardinally measurable, like income. Here we focus on the case where the variable is only ordinal, the case where cardinal comparisons of the variable have no meaning.

Our analysis addresses the fundamental problem of providing welfare rankings and inequality rankings over distributions where the variables in question are purely ordinal, both unidimensional and multidimensional. To this end, we consider the case of two dimensions (although the results can be extended to an arbitrary number of dimensions). We provide a complete methodology for how univariate and multivariate comparisons of inequality and welfare should be made for ordinal data. We define elementary transformations that reflect welfare increase and inequality reduction, empirically implementable criteria and classes of welfare functions that embody ethical concerns about welfare and inequality, and state a set of equivalence results. The results are clear and concise

[1]Inequality and risk comparisons share a common formal structure (Gajdos and Weymark, 2012). Atkinson (1970) compares income distributions in terms of their inequality, which is formally equivalent to what Rothschild and Stiglitz (1970) do for random variables. Here we concentrate on inequality.

[2]The foundation of these results was first established by Hardy et al. (1934) and then further developed by Kolm (1969), Atkinson (1970), Rothschild and Stiglitz (1970), Dasgupta et al. (1973), Sen (1973) and Fields and Fei (1978). Elementary transfers here are mean-preserving spreads that express the notion of inequality or risk increase.

[3]See, for example,Lehmann (1955), Kamae et al. (1977), Atkinson and Bourguignon (1982), Maasoumi (1986), Rothschild and Stiglitz (1970), Tsui (1995, 1999) and Gajdos and Weymark (2005).

owing to a unified framework which is integration on partially ordered sets.

The paper is organised as follows. Section 2 provides a general overview of the problem, a brief introduction of its many applications and a detailed description of the results we obtain. Section 3 sets out basic notation. Sections 4 and 5 contain the definitions and the results in the case of the one-dimensional and bidimensional distributions, respectively.

2 The background to the problem

The problem we face is that of making either welfare or inequality comparisons of something for which there are, loosely speaking, no numbers. The only information we have is about an underlying ordering of categories.

There are many applications of this general setting. Often a decision maker has to evaluate commodities according to criteria that are not inherently quantifiable; for example, when comparing two job offers, apart from a salary, a person takes into account the working environment, career development, the company's facilities and so on. In recent years ordinal indicators have been used extensively to assess nation's wellbeing and progress. Several countries and international organisations have launched initiatives to incorporate a multidimensional perspective of progress including education, health, environment, safety, governance and life satisfaction . Examples include the OECD's Better Life Index, the report by the commission on the measurement of economic performance and social progress (Stiglitz et al., 2009), the report by the Office for National Statistics on measuring national well-being in the UK. In health economics ordinal indicators of health are widespread, especially as survey data become more widely available and researchers try to measure health inequality using such data (Allison and Foster 2004, Apouey 2007, Abul Naga and Yalcin 2008, Kobus and Miłoś 2012, Apouey and Silber 2013, Lazar and Silber 2013, Abul Naga and Stapenhurst 2015, Gravel et al. 2014, Lv et al. 2015, Kobus 2015, Cowell and Flachaire 2017). Bond and Lang (2013, 2014) point out serious problems in educational economics and happiness economics when test scores or life satisfaction are treated as ordinal variables, as they should be. The same situation arises when one analyses, for example, the consumption of public goods such as quality of schooling. In risk analysis for credit scoring or questions to retail investors like mandatory risk tolerance survey by a retail brokerage company ordinal variables appear a lot.

Furthermore, bivariate distributions of ordinal indicators are often considered while measuring socioeconomic inequalities in health. Numerous studies in economics, social epidemiology, medicine and health psychology, medical sociology analyse socioeconomic inequalities in health (Deaton and Paxson 1998, Beckett 2000, Lynch 2003, Case and Deaton 2005, Herd 2006, Cutler et al. 2015, Gonzalez et al. 2016; see also Smith 1999, O'Donnell et al. 2008, Cutler et al. 2011, Evans et al. 2012 and Stowasser et al. 2014 for reviews). Generally, the lower the socioeconomic status (SES) the lower the health status (Adler et al., 1994). This phenomenon is observed in many industrialised countries and is known as the SES-health gradient (Marmot, 2006). There is evidence for developing countries as well (Strauss and Thomas, 1998). In empirical applications the SES is often proxied by income classes, educational attainment, social or occupational class and health is proxied by self-reported health status; thus bivariate distributions of ordinal indicators are analysed. Finally, whenever one wants to make comparisons that are robust to the coarsening of the support (for example where income is available only as a bracket variable in \$500 or \$1000 brackets, or where life expectancy data are available in brackets of 5 or 10 years), the current framework is appropriate.

2.1 An illustration

With ordinal data, standard notions of stochastic dominance are not well-defined, because classic notions of integration depends on the scale of the integrand. This problem applies to second and higher order of dominance. It does not apply to first order dominance because of scale invariance of the cumulative distribution function, however, the framework we introduce starts with first-order dominance and proceeds to the second-order. Here we have many scales – indeed all scales that preserve the underlying ordering are equally good – so there are many stochastic dominance curves, one for each scale. To see why this causes problems, let us look at the following example for second-order stochastic dominance.

Let F, G be the following distribution functions

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{6} & 0 \leq x < 2 \\ \frac{5}{6} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases} \qquad G(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{3} & 0 \leq x < 2 \\ \frac{2}{3} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

and

$$\int^x F(y)\mathrm{d}y = \begin{cases} 0 & x < 0 \\ \frac{1}{6}x & 0 \le x < 2 \\ \frac{1}{3} + \frac{5}{6}(x-2) & 2 \le x < 3 \\ \frac{7}{6} + (x-3) & 3 \le x \end{cases} \quad \int^x G(y)\mathrm{d}y = \begin{cases} 0 & x < 0 \\ \frac{1}{3}x & 0 \le x < 2 \\ \frac{2}{3} + \frac{2}{3}(x-2) & 2 \le x < 3 \\ \frac{4}{3} + (x-3) & 3 \le x \end{cases}$$

It is easy to check that $\int^x F(y)\mathrm{d}y \le \int^x G(y)\mathrm{d}y$ for all x, but if we change the scale such that 3 becomes 6 we get

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{6} & 0 \le x < 2 \\ \frac{5}{6} & 2 \le x < 6 \\ 1 & 6 \le x \end{cases} \quad G(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{3} & 0 \le x < 2 \\ \frac{2}{3} & 2 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

and

$$\int^x F(y)\mathrm{d}y = \begin{cases} 0 & x < 0 \\ \frac{1}{6}x & 0 \le x < 2 \\ \frac{1}{3} + \frac{5}{6}(x-2) & 2 \le x < 6 \\ \frac{11}{3} + (x-6) & 6 \le x \end{cases} \quad \int^x G(y)\mathrm{d}y = \begin{cases} 0 & x < 0 \\ \frac{1}{3}x & 0 \le x < 2 \\ \frac{2}{3} + \frac{2}{3}(x-2) & 2 \le x < 6 \\ \frac{10}{3} + (x-6) & 6 \le x. \end{cases}$$

Second order dominance no longer holds.

2.2 Previous approaches

Similar examples of ranking reversals have been provided for the mean, variance and standard inequality measures (Abul Naga and Yalcin 2008, Lazar and Silber 2013, Kobus 2015). The key concepts of inequality and risk measurement theory are difficult to interpret in the context of ordinal data. For example, with the Pigou-Dalton Transfer Principle is undefined because the mean – and hence a mean-preserving change in the distribution – is undefined. Without further assumptions on the underlying variable (which is in fact an implicit way of cardinalising an ordinal indicator (Bond and Lang, 2014), the impact of a Pigou-Dalton transfer on health inequality remains indeterminate. One solution is to develop inequality measurement theory such that it deals directly with distributions, and not variables for which there are no numbers. Indeed, inequality measurement theory has been extended in this direction (Allison and Foster 2004, Apouey 2007, Abul Naga and Yalcin 2008, Kobus and Miłoś 2012, Apouey and Silber 2013, Lazar and Silber 2013, Gravel et al. 2014, Abul Naga and Stapenhurst 2015, Kobus 2015, Cowell and Flachaire 2017). The problem is with choosing a perfectly equal distribution (and measuring inequality as the deviation from it) since there are many perfectly equal distributions (i.e. each such distribution where the whole mass is concentrated in one category). This problem has been dealt so far by resorting to the notion of a perfectly unequal distribution. Various ideas have been proposed in this domain (Blair and Lacy 2000, Abul Naga and Yalcin 2008, Zheng 2008, Apouey and Silber 2013, Gravel et al. 2014). The most well-known is the notion of most polarised distribution being the most unequal and the dominance relation related to it proposed by Allison and Foster (2004), (henceforth AF). It is a partial ordering which reflects the concept of median-preserving spreads of probability mass. Abul Naga and Yalcin 2008, Lazar and Silber (2013) and Kobus (2015) characterise indices based on the AF relation. Kobus and Miłoś (2012) characterise decomposable indices consistent with AF. Abul Naga and Stapenhurst (2015) develop estimation for inequality indices for ordinal data. AF has been criticised for focusing more on polarisation than inequality (Zheng 2008, Kobus 2015). Cowell and Flachaire (2017) take a different route, by considering various reference points, not only the median, and develop statistical properties of the class of indices. Lv et al. (2015) propose a class of measures which is not purely ordinal as they take into account a distance between categories, but which reflects the fact that in reality partial information may be available on how big the differences between categories can be. Recently, Allanson (2017) has provided a methodology for measuring social inequalities in health. In such a case, Zheng (2011) uses a revised notion of Lorenz dominance, but the method of Allanson (2017) has the advantage of being complete, in that it provides conclusive rankings.

To date there are very few contributions for multivariate data. Makdissi and Yazbeck (2014) propose a counting approach but this requires transforming each health indicator into a binary variable losing substantial part of the information; Sonne-Schmidt et al. (2016) propose criteria for comparing distributions which are only tractable in the case of two binary indicators, so their applicability is very restricted. Duclos and Échevin (2011)) propose a robust method for measuring health-income gradient based on dominance conditions, when health is ordinal and income is continuous.

2.3 Our approach

This paper gives a complete methodology for comparing distributions of ordinal data in terms of welfare and inequality. We do this by characterising an ordinal indicator as either a totally or partially ordered set and then using the notion of integration on partially ordered sets (Rota 1964, Parker and Ram 1997). This makes the results relatively straightforward to state concisely and to interpret intuitively. Equipped with such mathematical framework, we are able to provide a complete set of standard results for the case of ordinal data. In particular, we consider both unidimensional and bidimensional setting and then in each of these settings our goal is to obtain the Hardy-Littlewood-Pólya (Hardy et al. (1934)) type of result, namely, equivalence between three notions: (i) the unanimity of utilitarian social planners (ii) an implementable criterion that allows one to compare distributions consistently with (i), and (iii) elementary transformations that reflect the notion of welfare and/or inequality reduction. For welfare comparisons, because of the scale invariance property of the cdf, standard notions of first-order dominance, both in a unidimensional and multidimensional setting are well-defined. We cover them here as well in order to treat welfare, inequality and higher-order comparisons for ordinal data in a complete coherent framework. In the unidimensional case we deal with the complete ordering of categories of an ordinal indicator, whereas in the multidimensional case, there is only a partial ordering i.e. we do not impose a priori which is better, high school education and fair health status or higher education and bad health status.

The results are described in detail in each section; we now mention them briefly. For unidimensional welfare comparisons we show the equivalence between upward shifts, non-decreasing welfare functions and generalized majorization which is the classical majorization (Marshall and Olkin 1979) without ordering of the elements of vectors that are compared (Theorem 1). Here vectors are treated as probability distributions and the ordering is given by the ordering of categories of an ordinal variable. In all results elementary transformations are represented as multiplication by a matrix from a given (defined in Theorems) subclass of lower triangular (sub)stochastic matrices. For unidimensional inequality comparisons, so called generalized majorization of second order is equivalent to concave welfare functions and spread-reducing transformations (Theorem 2). Our definition of concavity takes into account differences in adjacent categories. This is related to a special case of Hammond transfer ((Hammond, 1976)). (Gravel et al., 2014) use general Hammond transfers, so our set is a special case of theirs, and our class of concave functions is larger than theirs. Furthermore, we propose an implementable criterion for both pure inequality comparisons and joint inequality and welfare comparisons (Theorem 3). These are, respectively, generalized majorization of second order and generalized non-decreasing majorization of second order. That is, in an ordinal setting we make a distinction similar to the distinction between Lorenz dominance which allows to compare only distributions with the same mean and Generalized Lorenz dominance which relaxes the same mean assumption. This has not yet been done for ordinal data and it is one of the key facts in inequality measurement theory.

Hardy-Littlewood-Pólya is a powerful result but its multivariate extensions are hard to obtain. Here we provide them in a bidimensional setting that extends naturally to a multidimensional case. To this end we introduce a novel definition of transfer sensitivity (Shorrocks and Foster, 1987), namely, association sensitivity. It involves putting mass on the counter-diagonal in the lower end of the distribution combined with putting mass on the diagonal in the higher end of the distribution. The latter is an association-decreasing switch while the former is an association-increasing switch. A class of welfare functions that increases following such transfer reflects the judgment that inequality (bivariate inequality) down the distribution is more important. We prove the relevant equivalences for such transformations depending on whether inequality (Theorem 5) or both inequality and welfare are involved (Theorem 6).

3 Notation and definitions

We define a numerical representation of categories of ordinal variables $\mathbb{I} := \mathbf{I}_1 \times \mathbf{I}_2 = \{1, \ldots, n\} \times \{1, \ldots, m\}$ which is arbitrary as long as it preserves the ordering ($m = 1$ for unidimensional distribution). $\mathbf{I}_1, \mathbf{I}_2$ are totally ordered sets and $\mathbb{I}$ is endowed with the usual partial order: $(i, j) \precsim (i', j')$ if and only if $i \leq i'$ and $j \leq j'$ for all i and j. Throughout the paper $\mathbb{I}, n, m$ are fixed unless we explicitly state otherwise.

Now let f be a probability distribution on the set $\mathbb{I}$. We work with probability distributions for convenience.[4] By defining probability distribution f on $\mathbb{I}$ we make it independent of scale; that is, if there are two different scales with the same number of categories on each dimension, then $\mathbb{I}$ does not change and a given probability distribution can be related to both scales. Obviously we require

[4]This is just for convenience. Our principal results do not require scale independence and could be obtained using mass (absolute numbers) rather than probabilities or proportions

$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} = 1$ and for all $(i,j) \in \mathbb{I}$, $f_{ij} \geq 0$ and . We define marginal distributions f^1, f^2 on, respectively, $\mathbf{I}_1, \mathbf{I}_2$ by

$$f_i^1 := \sum_{j=1}^{m} f_{ij} \quad f_j^2 := \sum_{i=1}^{n} f_{ij} \tag{1}$$

and cumulative distributions by

$$F_i^1 := \sum_{k=1}^{i} f_k \quad F_j^2 := \sum_{l=1}^{j} f_l \tag{2}$$

Obviously, we have $\sum_{i=1}^{n} f_i = 1$ and for all $i \in \mathbf{I}_1$, $f_i \geq 0$ and similarly for f^2.

A multidimensional cumulative distribution function F at (i,j) equals

$$F_{ij} := \sum_{k=1}^{i}\sum_{l=1}^{j} f_{kl} \tag{3}$$

Obviously, we have

$$f_{ij} = F(i,j) - F(i,j-1) - F(i-1,j) + F(i-1,j-1).$$

For unidimensional distribution ($\mathbb{I} = \mathbf{I}_1 \times \{1\}$) we will write $F_i = F_i^1 = F_{i1}$.

Let $\mathcal{U}$ be a set of utility functions and $\mathcal{P}$ be a set of probability distributions on $\mathbb{I}$. Further, let $W : \mathcal{U} \times \mathcal{P} \mapsto \mathbb{R}$ denote a welfare function such that $W(u,F) = \Sigma_{i=1}^{n} u(i) f_i$ in the unidimensional case and $W(u,F) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} u(i,j) f_{ij}$ in the bidimensional case.

4 The unidimensional case: total ordering of the categories of an ordinal variable

4.1 Overview

For *welfare comparisons* (Section 4.2) we introduce the notion of generalised majorization as our implementable criterion. While in classical majorization (Marshall and Olkin, 1979) the elements of the vector are ordered decreasingly, here we take into account the underlying ordering of categories of an ordinal indicator. Similarly to classical majorization being related to bistochastic matrices (Kolm, 1977) , generalised majorization is related to lower triangular stochastic matrices ($LTSM$) (i.e. column stochastic matrices with zero elements above the diagonal) which express the notion of elementary welfare-enhancing transformations (ii). We show that each $LTSM$ matrix is generated by exchange matrices which reflect upward shifts of probability mass. This result was proved by Parker and Ram (1997) and we build upon their results in what follows. Generalised majorization and $LTSM$ matrices are equivalent to a class of non-decreasing functions (Theorem 1).

In the case of *inequality comparisons* (Section 4.3) there is a fundamental problem of giving a precise meaning to the concept of inequality reduction. In the discrete framework there are potentially many definitions of concavity, depending on which differences are taken into account. We choose the notion of concavity that relies on the differences in adjacent categories. We introduce the notion of generalised majorization of second order as our implementable criterion. This, similarly to second-order stochastic dominance in the standard framework, comes from a double integral. The $LTSM^2$ matrices are essentially $LTSM$ matrices transformed by the operations of integration and differentiation on a totally ordered set. They are also generated by exchange matrices. Here elementary exchanges reduce the spread by transferring probability mass *up* to an adjacent category in the lower end of the distribution and balancing this transfer in the higher end of the distribution with a transfer of probability mass *down* to an adjacent category. We prove the equivalence between concave functions, $LTSM^2$ matrices and generalised majorization of second order (Theorem 2). That these transfers happen between adjacent categories is directly related to our definition of concavity. Such transfers are a version of a Hammond transfer (Hammond, 1976). A Hammond transfer is a type of transfer which reduces the gap between two individuals being in different categories of an ordinal attribute irrespective of whether the loss experienced by one person equals the gain by another. The Pigou-Dalton Transfer is a special case of the Hammond transfer where *the same* amount of income that is taken from the richer is given to the poorer, that is, the gain equals the loss. Indeed, in the context of ordinal data, when the magnitude of gain and loss cannot be compared, the Hammond transfer appears to be "a highly plausible instances of ordinal inequality reduction" (Gravel et al., 2014). However, the Hammond transfers that Gravel et al. (2014) invoke take place between *arbitrary* categories, so our set of Hammond transfers is a special case of theirs. Thus, our class of equivalent concave functions is larger than theirs. If a given function is consistent with the transfers in (Gravel et al., 2014), then it is consistent with ours since ours is implied by theirs, but not the opposite.

What is important is that generalised majorization of second order is an operational criterion for inequality reduction transfers *only.* Later in the paper we deal with *joint unidimensional inequality and welfare comparisons* (Section 4.4). To this end we introduce generalised non-decreasing majorization of second order as our implementable criterion. This is a modification of the definition of generalised majorization of second order which allows us to compare distributions jointly in terms of first and second order dominance. We obtain a very intuitive interpretation of this majorization, that is, it is consistent with substochastic $LTSM$ matrices which further means that it is characterised by two types of elementary transfers, namely, transfers that reduce the spread (inequality) and transfers that move mass up to better categories (welfare). To our best knowledge, this is the first attempt at the joint treatment of inequality and welfare in the context of ordinal data. This sheds some light on the normative theory of measurement for ordinal data in the manner of Atkinson-Kolm-Sen. We state the equivalence result (Theorem 3).

4.2 Welfare comparisons in one dimension

Definition 1. ***Utilitarian unanimity (Non-decreasing functions of one variable)***
Distribution F is preferred to distribution G by all utilitarian maximisers whose utility function is non-decreasing i.e. for all $u \in \mathcal{U}_{nd}$, where $\mathcal{U}_{nd} := \{u \in \mathcal{U} | u(i_2) \geq u(i_1) \text{ for all } i_2 > i_1\}$.

Classical majorization compares partial sums constructed from vectors ordered decreasingly. Here, since vectors are probability distributions and the underlying ordering is given by the ordering of categories on which probability distributions exist, ordering probability vectors in a decreasing fashion does not make much sense. Therefore, we use the following definition of generalised majorization for which the ordering is taken as given.

Definition 2. *Generalised majorization*
For probability distributions f, g

$$f \precsim g \qquad \text{iff} \qquad \begin{cases} \Sigma_{i=1}^{k} f_i \leqslant \Sigma_{i=1}^{k} g_i & k = 1, 2, ..., (n-1) \\ \Sigma_{i=1}^{n} f_i = \Sigma_{i=1}^{n} g_i \end{cases}$$

The following definition will be useful in defining an implementable criterion. Vector majorization holds if a given inequality relation holds for each coordinate.

Definition 3. *Vector majorization*
For probability distributions f, g (or any vectors f, g from $\mathbb{R}^n \backslash \{0\}$, such that $\Sigma_{i=1}^{n} f_i = \Sigma_{i=1}^{n} g_i$)

$$f \leqslant_{vec} g \qquad \text{iff} \qquad f_i \leqslant g_i \qquad i = 1, 2, ..., n$$

We are now ready to state what is meant by integration on a totally ordered domain.

Definition 4. *Integration on a totally ordered set*
The integration matrix $\int$ *for a total ordering is a lower triangular* $n \times n$ *matrix consisting of only* 1*, i.e.* $\int = (\zeta_{ij})$ *where* $\zeta_{ij} = 0$ *if* $j > i$ *and* $\zeta_{ij} = 1$ *elsewhere.*

For example, for $n = 6$ we have

$$\int = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Similarly, one can define differentiation on a totally ordered set.

Definition 5. *Differentiation on a totally ordered set*
The differentiation matrix ∂ *for a total ordering is a lower triangular* $n \times n$ *matrix consisting of* 1 *on diagonal,* -1 *on neighbourhood elements below diagonal and 0 elsewhere, i.e.* $\partial = (\mu_{ij})$ *where* $\mu_{ij} = 1$ *if* $j = i$*,* $\mu_{ij} = -1$ *if* $j + 1 = i$ *and* $\mu_{ij} = 0$ *elsewhere.*

For example, for $n = 6$ we have

$$\partial = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Remark 1. $\partial = \int^{-1}$

Using $\int$ generalised majorization can be now presented as vector majorization.

Corollary 1. ***Implementable criterion (Generalised majorization)***
Let f, g be probability distributions. We have

$$f \precsim g \qquad \text{iff} \qquad \int f \leqslant_{vec} \int g$$

We will now present elementary transformations that reduce welfare.

Definition 6. *Lower Triangular Stochastic Matrix (LTSM)*
Matrix is column stochastic if its elements are all non-negative and all columns add up to 1. LTSM is column stochastic matrix which has only zero elements above diagonal.

For example, for $n = 6$ we have

$$\begin{pmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 1 \end{pmatrix}$$

Definition 7. ***Elementary transformations (LTSM majorization)***
Let f, g be probability distributions.

$$f \precsim_{LTSM} g \qquad \text{iff} \qquad f = Lg \qquad \text{for some LTSM } L$$

Theorem 1 establishes the equivalence between utilitarian unanimity implementable criterion (generalised majorization) and multiplication via an $LTSM$ matrix (elementary transformations).

Theorem 1. ***Utilitarian unanimity of non-decreasing functions $\iff$ Generalised majorization $\iff$ LTSM Majorization***

In the proof we show that each elementary transfer defines the following elementary exchange matrix.

Definition 8. *Exchange matrix $L_{pq}(\varepsilon)$ differs from identity only on $\{p, q\} \times \{p, q\}$*

$$\begin{array}{c} \begin{pmatrix} \dots & p & \dots & q & \dots \end{pmatrix} \\ \begin{pmatrix} \dots \\ p \\ \dots \\ q \\ \dots \end{pmatrix} \begin{pmatrix} \ddots & & & & \\ & 1-\varepsilon & & & \\ & & \ddots & & \\ & \varepsilon & & 1 & \\ & & & & \ddots \end{pmatrix} \end{array}$$

where p and q indicate column row numbers.

Exchange matrices are obviously $LTSM$ matrices. They reflect elementary increments of ϵ of probability mass, or in other words, elementary welfare reduction transfers. Furthermore, the multiplication of two exchange matrices generates an $LTSM$ matrix which is a consequence of the following lemma.

Lemma 1. *LTSM matrices form a semigroup.*

As a consequence of the proof of Theorem 1 (see Appendix) we have the following lemma.

Lemma 2. *Each LTSM can be decomposed into a product of at most $n-1$ exchange matrices. LTSM semigroup is generated by exchange matrices.*

A semigroup is a a set of matrices closed under multiplication that includes identity matrix. Such property ensures that $LTSM$ majorization is a preorder i.e. the identity matrix ensures reflexivity and closure under multiplication ensures transitivity. Moreover, it is a partial order, because it is anti-symmetric e.g. if $f = L_1 g$ and $g = L_2 f$, then $f = L_1 L_2 f$, and L_1, L_2 have to be identity matrices because $LTSM$ matrices have only non-negative elements.

4.3 Inequality comparisons in one dimension

As mentioned in Section 2, in the definition of concavity we focus on adjacent categories.

Definition 9. ***Utilitarian unanimity (Concave functions of one variable)***
Distribution F is preferred to distribution G by all utilitarian maximisers whose utility function is

concave i.e. for all $u \in \mathcal{U}_{nd}$*, where*
$\mathcal{U}_c := \{u \in \mathcal{U} | u(i_1+1) - u(i_1) \geq u(i_2+1) - u(i_2) \text{ for all } i_1 < i_2\}$.

Similarly to the first order, we define generalised majorization of the 2nd order. As we will show further (Theorem 2), this criterion is an operation criterion that reflects *pure* inequality comparisons in an ordinal setting.

Definition 10. *Generalised majorization of 2nd order*
For probability distributions f, g

$$f \precsim^2 g \qquad \text{iff} \qquad \begin{cases} \Sigma_{j=1}^{l} \Sigma_{i=1}^{j} f_i \leqslant \Sigma_{j=1}^{l} \Sigma_{i=1}^{j} g_i & l = 1, 2, ..., (n-1) \\ \Sigma_{j=1}^{n} \Sigma_{i=1}^{j} f_i = \Sigma_{j=1}^{n} \Sigma_{i=1}^{j} g_i \end{cases}$$

Corollary 2. ***Implementable criterion (Generalised majorization of 2nd order)***
Let f, g *be probability distributions. We have*

$$f \precsim^2 g \quad \text{iff} \quad \int^2 f \leqslant_{vec} \int^2 g$$

We also introduce second order $LTSM$ majorization called $LTSM^2$ majorization. Matrices that belong to $LTSM^2$ are $LTSM$ matrices transformed by integration and differentiation matrices at the first order.

Definition 11. ***Elementary transformations (*$LTSM^2$ *majorization)***
Let f, g *be probability distributions.*

$$f \precsim_{LTSM^2} g \qquad \text{iff} \qquad f = L^2 g, \qquad \text{where} \qquad L^2 = \partial L \int \qquad \text{for some LTSM } L$$

Clearly, L^2 matrices form a semigroup.

Lemma 3. $LTSM^2$ *matrices form a semigroup.*

We have the following equivalence between utilitarian unanimity, majorization of the second order and inequality reducing transformations.

Theorem 2. ***Utilitarian unanimity of concave functions*** $\iff$ ***Generalised majorization of*** 2***nd order*** $\iff$ $LTSM^2$ ***Majorization***

Lemma 4. *Each* $L^2 \in LTSM^2$ *can be decomposed into at most* $n-1$ *matrices* $L^2_{pq}(\varepsilon)$ *exchange matrices (Definition 12).*

Definition 12. $L^2_{pq}(\varepsilon)$ *exchange matrix*

$$L^2_{pq}(\varepsilon) = \partial L_{pq}(\varepsilon)\int = Id + \begin{matrix} \vdots \\ p-1 \\ p \\ p+1 \\ p+2 \\ \vdots \end{matrix} \begin{pmatrix} \vdots \\ 0 \\ -1 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 1_p & 0_{p+1} & \dots & 0 \end{pmatrix} \varepsilon +$$

$$\begin{matrix} \vdots \\ q-1 \\ q \\ q+1 \\ q+2 \\ \vdots \end{matrix} \begin{pmatrix} \vdots \\ 0 \\ 1 \\ -1 \\ 0 \\ \vdots \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 1_p & 0_{p+1} & \dots & 0 \end{pmatrix} \varepsilon =$$

$$\begin{matrix} \vdots \\ p-1 \\ p \\ p+1 \\ p+2 \\ \vdots \\ q-1 \\ q \\ q+1 \\ q+2 \\ \vdots \end{matrix} \begin{pmatrix} \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1_{p-1} & 0 & 0 & 0 & \dots & 0 \\ -\varepsilon & -\varepsilon & \dots & -\varepsilon & (1-\varepsilon)_p & 0 & 0 & \dots & 0 \\ \varepsilon & \varepsilon & \dots & \varepsilon & \varepsilon & 1_{p+1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1_{p+2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \varepsilon & \varepsilon & \dots & \varepsilon & -\varepsilon & 0 & 0 & \dots & 0 \\ -\varepsilon & -\varepsilon & \dots & -\varepsilon & -\varepsilon & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

As we see, elementary exchange matrices for second order majorization reduce the spread of the distribution by transferring ϵ probability mass to an adjacent category up (from p to $p+1$) and down (from $q+1$ to q). That these categories are adjacent is related to our definition of concavity. The elementary exchanges are a special case of Hammond transfers (Hammond, 1976). Let us look at the following examples of exchange matrices.

For example, let $g=(g_1,\dots,g_9)$ be a probability distribution and $L_{36}(\varepsilon)$ be an exchange matrix $(p=3,q=6,n=9)$.

$$L_{36}(\varepsilon)=\begin{pmatrix}1&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&0&1-\varepsilon&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&\varepsilon&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&1\end{pmatrix}$$

We have $(L_{36}(\varepsilon)g)^T=(g_1,g_2,g_3-\varepsilon g_3,g_4,g_5,\varepsilon g_3+g_6,g_7,g_8,g_9)$ and $(\int L_{36}(\varepsilon)g)^T=(G_1,G_2,G_3-\varepsilon g_3,G_4-\varepsilon g_3,G_5-\varepsilon g_3,G_6,G_7,G_8,G_9)$. Further, let $L^2_{36}(\varepsilon)$ be an exchange matrix.

$$L^2_{36}(\varepsilon)=\begin{pmatrix}1&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\-\varepsilon&-\varepsilon&1-\varepsilon&0&0&0&0&0&0\\\varepsilon&\varepsilon&\varepsilon&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\\varepsilon&\varepsilon&\varepsilon&0&0&1&0&0&0\\-\varepsilon&-\varepsilon&-\varepsilon&0&0&0&1&0&0\\0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&1\end{pmatrix}$$

We have $(L^2_{36}(\varepsilon)g)^T=(g_1,g_2,g_3-\varepsilon(g_1+g_2+g_3),\varepsilon(g_1+g_2+g_3)+g_4,g_5,\varepsilon(g_1+g_2+g_3)+g_6,-\varepsilon(g_1+g_2+g_3)+g_7,g_8,g_9)$ and $(\int L^2_{36}(\varepsilon)g)^T=(G_1,G_2,G_3-\varepsilon G_3,G_4,G_5-\varepsilon G_3,G_6,G_7,G_8,G_9)$.

As another example, let $g=(g_1,\dots,g_9)$ be a probability distribution and $L:=L_{36}(\frac{1}{2})L_{56}(\frac{1}{4})$ be $LTSM$ matrix.

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We have $(L^2g)^T = \Big(g_1, g_2, g_3 - \frac{1}{2}G_3, g_4 + \frac{1}{2}G_3, g_5 - \frac{1}{4}G_5, g_6 + \frac{1}{2}G_5 + \frac{1}{2}G_3, g_7 - \frac{1}{4}G_5 - \frac{1}{2}G_3, g_8, g_9\Big)$. Furthermore, let $L^2 = L^2_{36}(\frac{1}{2})L^2_{56}(\frac{1}{4})$ be $LTSM^2$ matrix.

$$L^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 & 0 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We have $(\int L^2g)^T = L\int g = \Big(G_1, G_2, \frac{1}{2}G_3, G_4, \frac{3}{4}G_5, G_6 + \frac{1}{2}G_3 + \frac{1}{4}G_5, G_7, G_8, G_9\Big)$.

4.4 Joint welfare and inequality comparisons

Theorem 2 characterises concave utility ordering, however, one might be interested in joint inequality and welfare reduction. The appropriate class of functions then is non-decreasing and concave.

Definition 13. ***Utilitarian unanimity (Non-decreasing concave functions of one variable)***

Distribution F is preferred to distribution G by all utilitarian maximisers whose utility function is

non-decreasing and concave i.e. for all $u \in \mathcal{U}_{nd}$, *where*

$\mathcal{U}_{ndc} := \{u \in \mathcal{U} | u(i_1+1) - u(i_1) \geq u(i_2+1) - u(i_2) \geq 0$ *for all* $i_1 < i_2\}$.

Let us note that if $f \precsim g$ and $\Sigma_{j=1}^{n}\Sigma_{i=1}^{j} f_i = \Sigma_{j=1}^{n}\Sigma_{i=1}^{j} g_i$ then $f = g$, so $\precsim^2$ lets us compare only distributions that are not comparable via $\precsim$. By dropping $\Sigma_{j=1}^{n}\Sigma_{i=1}^{j} f_i = \Sigma_{j=1}^{n}\Sigma_{i=1}^{j} g_i$ from the definition of generalised majorization we can compare distribution both in terms of inequality and welfare reduction.

Corollary 3. ***Implementable criterion (Generalised nondecreasing majorization on 2nd order)***

For probability distributions f, g

$f \precsim_{nd}^2 g$ *iff* $\Sigma_{j=1}^{l}\Sigma_{i=1}^{j} f_i \leqslant \Sigma_{j=1}^{l}\Sigma_{i=1}^{j} g_i \qquad l = 1, 2, ..., n$

Let us now introduce a semigroup of **sub**stochastic lower triangular matrices ($sLTSM$):

$$L \in sLTSM \iff \forall_j \ \Sigma_{i=1}^{n} l_{ij} \leq 1 \ \wedge \ \forall_{i,j} \ l_{ij} \geq 0 \ \wedge \ \forall_{j>i} \ l_{ij} = 0$$

For example, for $n = 6$ we have

$$\begin{pmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ \frac{1}{6} & \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 1 \end{pmatrix}$$

Remark 2. *For every* $sLTSM$ *matrix* L *there exists* $LTSM$ *matrix* L' *such that* $L = L'D$ *for some diagonal matrix* D *with* $0 < d_{ii} \leq 1$ *for all* i.

Remark 3. $sLTSM$ *is generated by exchange matrices* $L_{pq}(\varepsilon)$ *and by subtraction matrices* $L_p^s(\varepsilon)$, *which differs from identity matrix only on element* $l_{pp}^s = 1 - \varepsilon$

For example,

$$\begin{pmatrix} 1-\varepsilon & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\varepsilon & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \varepsilon & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We are now ready to introduce

Definition 14. ***Elementary transformations (*$sLTSM^2$ *majorization)***
Let f, g *be probability distributions.*

$$f \precsim_{sLTSM^2} g \qquad \textit{iff} \qquad f = L^2 g, \qquad \textit{where} \qquad L^2 = \partial L \int \qquad \textit{for some sLTSM } L$$

In order to find the form of the elementary $sLTSM^2$ matrix we transform the exchange matrix and the subtraction matrix via $\partial - \int$. We already know the form of $L_{pq}(\varepsilon)$ matrices under this transformation (see Definition 12). The form of $L_p^s(\varepsilon)$ is the following.

$$L_p^{s\,2}(\varepsilon) = \partial L_p^s(\varepsilon) \int = Id + \begin{matrix} \vdots \\ p-1 \\ p \\ p+1 \\ p+2 \\ \vdots \end{matrix} \begin{pmatrix} \vdots \\ 0 \\ -1 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 1_p & 0_{p+1} & \dots & 0 \end{pmatrix} \varepsilon =$$

$$\begin{matrix} \vdots \\ p-1 \\ p \\ p+1 \\ p+2 \\ \vdots \end{matrix} \begin{pmatrix} \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1_{p-1} & 0 & 0 & 0 & \dots & 0 \\ -\varepsilon & -\varepsilon & \dots & -\varepsilon & (1-\varepsilon)_p & 0 & 0 & \dots & 0 \\ \varepsilon & \varepsilon & \dots & \varepsilon & \varepsilon & 1_{p+1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1_{p+2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Theorem 3. ***Utilitarian unanimity of non-decreasing concave functions*** $\iff$ ***Generalised nondecreasing majorization of*** $2nd$ ***order*** $\iff$ $sLTSM^2$ ***Majorization***

5 The bidimensional case: partial ordering of categories of two ordinal variables

5.1 Overview

For *welfare comparisons* distributions are defined on a partially ordered set and the notion of generalised majorization accounts for the fact that some elements are incomparable: there are zero entries in the integral matrix indicating lack of comparison. Furthermore, similarly to standard framework, there are different multivariate generalisations of unidimensional welfare comparisons. We start with first order stochastic dominance (FSD) of the Atkinson-Bourguignon type (Atkinson and Bourguignon 1982, 1987), which is called bidimensional generalised orthant majorization. It allows for upward transfers of probability mass along the marginals and association-increasing switches, thus it is equivalent to non-decreasing and submodular welfare functions. The multivariate counterpart to $LTSM$ matrices, namely $\widetilde{LTSM}$ matrices are generated by exchange matrices such that an exchange of probability mass takes place only between comparable categories; hence these matrices are a proper subset of $LTSM$ matrices. They imply but are not equivalent to bidimensional generalised orthant majorization (Theorem 4). That is, there are distributions which fulfil bidimensional generalised orthant majorization, but for which $\widetilde{LTSM}$ matrix does not exist. Therefore, we study an implementable criterion which is equivalent to the set of $\widetilde{LTSM}$ matrices and this is first order dominance (FOD) (Lehmann 1955, Levhari et al. 1975, Østerdal 2010) which we call bidimensional generalised majorization. Here integration takes places on so called comprehensive sets (Kamae et al. 1977, Mosler and Scarsini 1991). Only upward transfers of probability mass are allowed and thus the equivalent class of welfare functions is larger than in the case of FSD and contains non-decreasing functions. We prove the equivalence between this class, $\widetilde{LTSM}$ matrices and bidimensional generalised majorization (Theorem 4), that is, we repeat the standard result concerning FOD dominance taking into account that we integrate on a partially ordered set.

For *inequality comparisons* we concentrate on second-order generalisations of FOD. We introduce the notion of bidimensional generalised majorization of the second order which is an integration on comprehensive sets of *cumulative* probability vectors. We introduce a novel definition of transfer sensitivity in the bivariate context, namely, association sensitivity. We obtain the result which is very interesting not only in an ordinal context, but in general, for the characterisation of multivariate dominance orderings. It is well-known that multivariate generalisations of the famous Hardy-Littlewood-Polya result (Trannoy 2006, Gravel and Moyes 2012) are difficult to obtain in a multivariate setting. As Gravel and Moyes (2012) point out the equivalence is typically established by utilitarian unanimity and *either* an implementable criterion (Atkinson and Bourguignon 1982, Decancq 2012) *or* elementary transformations (Kolm 1977, Müller and Scarsini 2012), but not amongst all three. We provide such an equivalence result (Theorem 6). In particular, bidimensional generalised majorization of second order is equivalent to association - sensitivity. The latter are transfers in which we reduce association in the lower tail by transferring mass to the counter diagonal and at the same time we increase association in the upper tail by transferring mass to the diagonal. In other words, a correlation decreasing switch in the lower tail is followed by a correlation increasing switch in the upper tail). Since reducing bivariate inequality is more important lower down the distribution, the net effect of such a composite transfer is inequality reducing and welfare increasing. This is similar in spirit to Transfer Sensitivity Principle (Shorrocks and Foster, 1987), which in a unidimensional cardinal setting states that inequality is reduced following a progressive combined with a regressive transfer at a higher income level. Here, we exchange association. We show (Theorem 6) that the sequence of such transfers is equivalent to bidimensional generalised majorization of second order and to a class of non-decreasing, submodular functions that are increasing with respect to our bivariate association - exchanging transfers.

5.2 Welfare comparisons in two dimensions

Definition 15. ***Utilitarian unanimity (Non-decreasing and submodular functions of two variables)***

Distribution F is preferred to distribution G by all utilitarian maximisers whose utility function is

non-decreasing and submodular i.e. for all $u \in \mathcal{U}_{ndsub}$*, where*

$$\mathcal{U}_{ndsub} := \{u \in \mathcal{U} | u(i_2, j) \geq u(i_1, j),\ u(i, j_2) \geq u(i, j_1) \text{ and }$$
$$u(i_2, j_2) - u(i_2, j_1) - u(i_1, j_2) + u(i_1, j_1) \leq 0 \text{ for all } i_2 > i_1,\ j_2 > j_1\}$$

We complement the partial order $\precsim$ on the set $\mathbb{I}$ to a linear order in an arbitrary way which is consistent with the partial order. In other words, we numerate the elements of $\mathbb{I}$, such that for any $(i_k, j_k), (i_l, j_l) \in \mathbb{I}$ if $(i_k, j_k) \precsim (i_l, j_l)$ then $k \leq l$.

Definition 16. *Bidimensional generalised orthant majorization*
For probability distributions f, g

$$f \precsim g \qquad \text{iff} \qquad \begin{cases} \Sigma_{i=1}^{k}\Sigma_{j=1}^{l} f_{ij} \leqslant \Sigma_{i=1}^{k}\Sigma_{j=1}^{l} g_{ij} & k = 1, 2, \ldots, (n-1) \quad l = 1, 2, \ldots, (m-1) \\ \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} f_{ij} = \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} g_{ij} & \end{cases}$$

We now deal with a partial order and therefore the notion of integration has to be extended appropriately.

Definition 17. *Integration on a partially ordered set*
Integration matrix $\int$ *for a partial ordering is a lower triangular matrix, where* $\zeta_{ij} = 1$ *if* $j \precsim i$ *for* $i, j \in \mathbb{I}$ *and* $\zeta_{ij} = 0$ *elsewhere.*

For example, for $\mathbb{I} = \{1, 2\} \times \{1, 2\}$, elements $(1, 1)$ and $(2, 2)$ comparable, but $(1, 2)$ and $(2, 1)$ are not. Non-comparable elements are denoted as zeros in the integration matrix.

$$\int = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

while the Mobius inversion matrix is

$$\partial = \int^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 \end{pmatrix}$$

Similarly to a unidimensional case, by using $\int$ we can present bidimensional generalised majorization as vector majorization.

Corollary 4. ***Implementable criterion (Bidimensional generalised orthant majorization)***
Let f, g be probability distributions. We have

$$f \precsim g \qquad \text{iff} \qquad \int f \leqslant_{vec} \int g$$

Given that some elements are incomparable in a partial order, we introduce exchange matrices such that a transfer only takes place if $p \precsim q$. $\widetilde{LTSM}$ is a subset of $LTSM$ matrices that are generated by such exchange matrices.

Definition 18. ***Elementary transformations ($\widetilde{LTSM}$ majorization)***
For a given partial ordering $\widetilde{LTSM}$ is an LTSM matrix that belongs to the following semigroup

$$\widetilde{LTSM} = \langle L_{pq}(\varepsilon) | p \precsim q, \varepsilon \in [0, 1) \rangle$$

Lemma 5. *For a given partial ordering $\widetilde{LTSM}$ consists of $LTSM$ matrices such that if $\zeta_{ij} = 0$ then $l_{ij} = 0$.*

Theorem 4. ***$\widetilde{LTSM}$ Majorization $\Rightarrow$ Bidimensional generalised orthant majorization $\iff$ Utilitarian unanimity of non-decreasing and submodular functions***

The implication is only in one direction for the following reason. Let us consider two distributions on $\{1, 2\} \times \{1, 2\}$:

$$x_{11} = 0.2 \;\; x_{12} = 0.2 \;\; x_{21} = 0.2 \;\; x_{22} = 0.4$$

$$y_{11} = 0.5 \;\; y_{12} = 0 \;\; y_{21} = 0 \;\; y_{22} = 0.5$$

Then of course we have $\int x <_{vec} \int y$, but $\widetilde{LTSM}$ matrix L such that $x = Ly$ would have to meet condition

$$l_{41} \cdot 0.5 + l_{44} \cdot 0.5 = 0.4.$$

However, we know that $l_{44} = 1$ because L is $LTSM$ and we calculate that $l_{41} = -0.6$ so such $\widetilde{LTSM}$ matrix does not exist.

Our goal now is to come up with a criterion that ensures equivalence and the following is related to the literature on FOD dominance.

Definition 19. ***Utilitarian unanimity (Non-decreasing functions of two variables)***
Distribution F is preferred to distribution G by all utilitarian maximisers whose utility function is non-decreasing, where

$$\mathcal{U}_{nd} := \{u \in \mathcal{U} | u(i_2, j) \geq u(i_1, j) \text{ and } u(i, j_2) \geq u(i, j_1) \text{ for all } i_2 > i_1,\ j_2 > j_1\}$$

We need a stronger condition than $\int x <_{vec} \int y$, therefore we define

Definition 20. *Comprehensive set*
$K \subseteq \mathbb{I}$ is comprehensive set if for any $(i_k, j_k) \in K$ we have $(i_l, j_l) \in K$ for any $(i_l, j_l) \precsim (i_k, j_k)$

Definition 21. *Bidimensional generalised majorization*

$$f \precsim g \qquad \text{iff} \qquad \Sigma_{(i,j)\in K} g_{ij} \leq \Sigma_{(i,j)\in K} f_{ij} \qquad \text{for any comprehensive set} \qquad K \subseteq \mathbb{I}$$

There are $\binom{n+m}{m} - 1$ comprehensive sets in $\mathbb{I}$. Let us consider the set $\tilde{\mathbb{I}} = (\{0\} \cup \mathbf{I}_1) \times (\{0\} \cup \mathbf{I}_2)$. Number of comprehensive sets in $\mathbb{I}$ is the number of paths from $(n, 0)$ to $(0, m)$ in $\tilde{\mathbb{I}}$ such that we can move $(-1, 0)$ or $(0, 1)$ in each step, neglecting the path that goes through point $(0, 0)$. Let us number those sets in a way that if $K_j \subset K_i$ then $j < i$.

Definition 22. *Integration on comprehensive sets*
We define $\widetilde{\int}$ as $\left[\binom{n+m}{m} - 1\right] \times nm$ matrix with $\tilde{\zeta}_{ij} = 1$ if for $j \in \mathbb{I}$ and for $i \leq \binom{n+m}{m} - 1$ we have $\tilde{\zeta}_{ij} = 1$ if $j \in K_i$, and $\tilde{\zeta}_{ij} = 0$ elsewhere.

Corollary 5. ***Implementable criterion (Bidimensional generalised majorization)***
Let f, g be probability distributions. We have

$$f \precsim g \qquad \text{iff} \qquad \widetilde{\int} f \leqslant_{vec} \widetilde{\int} g$$

We will now prove a Lemma which will be useful in the proof of equivalence Theorem.

Lemma 6. *If distributions f and g have the same marginal distributions and $\tilde{\int} f \leq_{vec} \tilde{\int} g$ then $f = g$.*

Theorem 5. ***Utilitarian unanimity of non-decreasing functions*** $\iff$ ***Bidimensional generalised majorization*** $\iff$ $\widetilde{LTSM}$ ***Majorization***

5.3 Joint welfare and inequality comparisons

We define the following utility function u:

$$v(i,j) := \begin{cases} u(i,j) - u(i,j+1) - u(i+1,j) + u(i+1,j+1) & \text{for } i < n,\, j < m \\ u(n,j) - u(n,j+1) & \text{for } i = n,\, j < m \\ u(i,m) - u(i+1,m) & \text{for } i < n,\, j = m \\ u(n,m) & \text{for } i = n,\, j = m \end{cases}.$$

Definition 23. ***Utilitarian unanimity***
Distribution F is preferred to distribution G by all utilitarian maximisers whose utility functions belong to the set

$$\mathcal{U}_{nd2} : v(i+1,j) - v(i,j) \geq 0 \textit{ and } v(i,j+1) - v(i,j) \geq 0 \textit{ for all } i,j\}$$

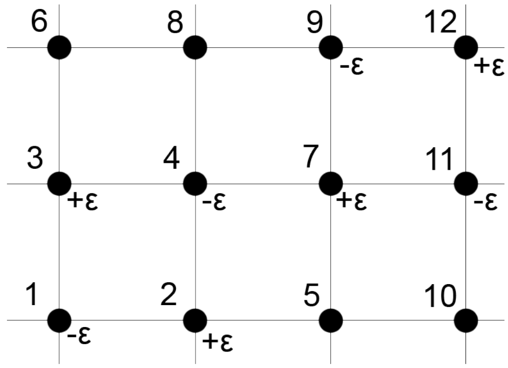
Less formally, functions that belong to $\mathcal{U}_{nd2}$ are non-decreasing, submodular and association-sensitive i.e. they increase following a transfer presented in Figure 1.

Definition 24. $\widetilde{sLTSM}^2$ *majorization*
$\widetilde{sLTSM}^2 = \langle \partial L_1 \int,\, L_2 | L_1 \in \widetilde{sLTSM}, L_2 \in \widetilde{LTSM} \rangle$

Let us take a closer look at the form of exchange matrices under $\partial - \int$ transformation. Figure 1 presents an example of such a matrix, $L^2_{17}(\varepsilon)$.

We transfer mass from the first category to the seventh category and then we compensate this transfer on categories that directly follow first and seventh category according to an underlying ordering - namely second and third after first and ninth and eleventh after seventh. This way we made an association-decreasing transfer in the bottom of the distribution and we compensated it with the association-increasing transfer at the top of the distribution.

Figure 1: Association - exchanging transfers



Definition 25. *Bidimensional generalised majorization of 2nd order*

$f \precsim^2 g$ *iff* $\Sigma_{(i,j)\in K}\Sigma_{(k,l)\precsim(i,j)}g_{kl} \leq \Sigma_{(i,j)\in K}\Sigma_{(k,l)\precsim(i,j)}f_{kl}$ *for any comprehensive set* $K \subseteq \mathbb{I}$

If we added the condition of sum equalisation on comprehensive sets, similarly to Definition 10, we would deal with bidimensional pure inequality comparisons. This follows Section 4.3 and this Section, so we decided to skip this.

Corollary 6. ***Implementable criterion (Bidimensional generalised majorization on 2nd order)***
Let f, g be probability distributions. We have

$$f \precsim^2 g \quad \textit{iff} \quad \widetilde{\int}\int f \leqslant_{vec} \widetilde{\int}\int g$$

Theorem 6. $\widetilde{sLTSM}^2$ ***Majorization*** $\Leftrightarrow$ ***Bidimensional generalised majorization of 2nd order*** $\Leftrightarrow$ ***Utilitarian unanimity***

6 Conclusion

This paper provides a complete set of results for measuring welfare and inequality for ordinal data. The results we obtain are the following. For unidimensional welfare comparisons, the relevant notion of elementary welfare-improving transformations are $LTSM$ matrices which describe upward shifts of probability mass and are thus equivalent to non-decreasing welfare functions. For unidimensional inequality comparisons, the relevant notion of elementary inequality-reducing transformation are $LTSM^2$ matrices which describe the reduction of spread and are thus equivalent to concave welfare functions, where the definition of concavity we analyse is concerned with transfers in adjacent categories. For unidimensional joint welfare and inequality comparisons the relevant notion of elementary welfare-improving and inequality-reducing transformation are $sLTSM^2$ matrices which describe transfers that move mass upwards (enhance welfare) and reduce spread (reduce inequality) and are thus equivalent to non-decreasing concave welfare functions.

Multivariate analogues of the Hardy et al. (1934) result are lacking in the literature (Gravel and Moyes, 2012). This paper fills that gap. For bidimensional welfare comparisons, we obtain equivalence equivalence results by introducing a novel definition of transfer sensitivity (Shorrocks and Foster, 1987), namely, association sensitivity. It involves a composite transfer which combines association-decreasing switch with an association-increasing switch at higher categories. A decision maker who prefers more reducing multivariate inequality lower down the distribution favours such a transfer. Formally, the relevant notion of elementary welfare-improving transformation are $\widetilde{sLTSM}^2$ matrices which describe upward shifts movements where it is feasible, namely, where elements are comparable. This is equivalent to non-decreasing functions. For bidimensional inequality and welfare comparisons, the relevant notion of elementary welfare-improving and inequality-reducing transformation are $\widetilde{sLTSM}^2$ matrices which describe the relevant transfers that move mass upwards, decrease association and combine association-decrease with association-increase at higher categories. These transformations are equivalent to a class of functions which are non-decreasing, submodular and association-sensitive.

References

Abul Naga, R. H. and C. Stapenhurst (2015). Estimation of inequality indices of the cumulative distribution function. *Economics Letters 130*, 109–112.

Abul Naga, R. H. and T. Yalcin (2008). Inequality measurement for ordered response health data. *Journal of Health Economics 27*, 1614–1625.

Adler, N., T. Boyce, and M. Chesney (1994). Socioeconomic status and health: The challenge of the gradient. *The American Psychologist 49*, 15–24.

Allanson, P. (2017). Monitoring income-related health differences between regions in great britain: A new measure for ordinal health data. *Social Science & Medicine 175*, 72–80.

Allison, R. A. and J. E. Foster (2004). Measuring health inequality using qualitative data. *Journal of Health Economics 23*, 505–552.

Apouey, B. (2007). Measuring health polarization with self-assessed health data. *Health Economics 16*(9), 875–894.

Apouey, B. H. and J. Silber (2013). Inequality and bi-polarization in socioeconomic status and health: Ordinal approaches. *Research on Economic Inequality 21*.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory 2*, 244–263.

Atkinson, A. B. and F. Bourguignon (1982). The comparison of multi-dimensional distributions of economic status. *Review of Economic Studies 49*, 183–201.

Atkinson, A. B. and F. Bourguignon (1987). Income distribution and differences in needs. In G. R. Feiwel (Ed.), *Arrow and the foundations of the theory of economic policy*, Chapter 12, pp. 350–370. New York: Macmillan.

Beckett, M. (2000). Converging health inequalities in later life–an artifact of mortality selection? *Journal of Health & Social Behavior 41*(1), 106–119.

Blair, J. and M. G. Lacy (2000). Statistics of ordinal variation. *Sociological methods and research 28*, 251–279.

Bond, T. and K. Lang (2013). The evolution of the black-white test score gap in grades K-3: The fragility of results. *The Review of Economics and Statistics 95*, 1468–1479.

Bond, T. and K. Lang (2014). The sad truth about happiness scales. Technical Report NBER Working Paper No. 19950, National Bureau of Economic Research, http://www.nber.org/papers/w19950.

Case, A. and A. Deaton (2005). Broken down by work and sex: how our health declines. In *Analyses in the economics of aging* (Wise, D. A. ed.)., pp. 185–205. Chicago. Chicago University Press.

Cowell, F. A. and E. Flachaire (2017). Inequality with ordinal data. *Economica forthcoming.*

Cutler, D. M., S. Huang, and A. Lleras-Muney (2015). When does education matter? the protective effect of education for cohorts graduating in bad times. *Social Science and Medicine 127*, 63–73.

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2011). Socioeconomic status and health: Dimensions and mechanisms. In *The Oxford Handbook of Health Economics*, Number NBER Working Paper No. 14333. http://www.nber.org.gate2.library.lse.ac.uk/papers/w14333: Oxford University Press.

Dasgupta, P. S., A. K. Sen, and D. A. Starrett (1973). Notes on the measurement of inequality. *Journal of Economic Theory 6*, 180–187.

Deaton, A. and C. Paxson (1998). Aging and inequality in income and health. *American Economic Review 88*, 248–253.

Decancq, K. (2012). Elementary multivariate rearrangements and stochastic dominance on a fréchet class. *Journal of Economic Theory 147*, 1450–1459.

Duclos, J.-Y. and D. Échevin (2011). Health and income: A robust comparison of Canada and the US. *Journal of Health Economics 30*, 293–302.

Evans, W., B. Wolfe, and N. Adler (2012). The SES-health gradient: a brief review of the literature. In W. E. B. Wolfe and T. Seeman (Eds.), *The Biological Consequences of Socioeconomic Inequalities*, Chapter The SES-health gradient: a brief review of the literature. Russell Sage Foundation.

Fields, G. S. and J. C. H. Fei (1978). On inequality comparisons. *Econometrica 46*, 303–316.

Gajdos, T. and J. Weymark (2005). Multidimensional generalized Gini indices. *Economic Theory 26*, 471–496.

Gajdos, T. and J. A. Weymark (2012). Introduction to inequality and risk. *Journal of Economic Theory 147*, 1313–1330.

Gonzalez, M., D. Swanson, M. Lynch, and G. Williams (2016). Testing satisfaction of basic psychological needs as a mediator of the relationship between socioeconomic status and physical and mental health. *Journal of Health Psychology 21*(6), 972–982.

Gravel, N., B. Magdalou, and P. Moyes (2014). Ranking distributions of an ordinal attribute. AMSE working paper 2014-50, Aix-Marseille School of Economics.

Gravel, N. and P. Moyes (2012). Ethically robust comparisons of bidimensional distributions with an ordinal attribute. *Journal of Economic Theory 147*, 1384–1426.

Hammond, P. J. (1976). Equity in two person situations: Some consequences. *Econometrica 47*, 1127–1135.

Hardy, G., J. Littlewood, and G. Pólya (1934). *Inequalities.* London: Cambridge University Press.

Herd, P. (2006). Do functional health inequalities decrease in old age?: Educational status and functional decline among the 1931-1941 birth cohort. *Research on Aging 28*(3), 375–392.

Kamae, T., U. Krengel, and G. L. O'Brien (1977). Stochastic inequalities on partially ordered spaces. *Annals of Probability 5*.

Kobus, M. (2015). Polarization measurement for ordinal data. *Journal of Economic Inequality 13*, 275–297.

Kobus, M. and P. Miłoś (2012). Inequality decomposition by population subgroups for ordinal data. *Journal of Health Economics 31*, 15–21.

Kolm, S.-C. (1969). The optimal production of social justice. In J. Margolis and H. Guitton (Eds.), *Public Economics*, pp. 145–200. London: Macmillan.

Kolm, S.-C. (1977). Multidimensional egalitarianisms. *Quarterly Journal of Economics 91*, 1–13.

Lazar, A. and J. Silber (2013). On the cardinal measurement of health inequality when only ordinal information is available on individual health status. *Health Economics 22*, 106–113.

Lehmann, E. (1955). Ordered families of distributions. *The Annals of Mathematical Statistics 26*, 399–419.

Levhari, D., J. Paroush, and B. Peleg (1975). Efficiency analysis for multivariate distributions. *Review of Economic Studies 42*, 87–91.

Lv, G., Y. Wang, and Y. Xu (2015). On a new class of measures for health inequality based on ordinal data. *Journal of Economic Inequality 13*, 465–477.

Lynch, S. (2003). Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. *Demography 40*(2), 309–331.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica 54*, 991–997.

Makdissi, P. and M. Yazbeck (2014). Measuring socioeconomic health inequalities in presence of multiple categorical information. *Journal of Health Economics 34*, 84–95.

Marmot, M. (2006). Status syndrome: A challenge to medicine. *Journal of the American Medical Association 295*(11), 1304–07.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory and Majorization*. New York: Academic Press.

Mosler, K. and M. Scarsini (1991). Stochastic orders and decision under risk. Technical report, Institute of Mathematical Statistics, Heyward, California.

Müller, A. and M. Scarsini (2012). Fear of loss, inframodularity, and transfers. *Journal of Economic Theory 147*, 1490–1500.

O'Donnell, O., E. van Doorslaer, A. Wagstaff, and M. Lindelow (2008). Analyzing health equity using household survey data. Learning resources series, World Bank Institute WBI.

Østerdal, L. (2010). The mass transfer approach to multivariate discrete first order stochastic dominance: direct proof and implications. *Journal of Mathematical Economics 46*, 1222–1228.

Parker, D. S. and P. Ram (1997). Greed and majorization. Technical report, Department of Computer Science, University of California, Los Angeles.

Rota, G.-C. (1964). On the foundations of combinatorial theory I. Theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2*, 340–368.

Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. A definition. *Journal of Economic Theory 2*, 225–243.

Sen, A. K. (1973). *On Economic Inequality.* Oxford: Clarendon Press.

Shorrocks, A. F. and J. E. Foster (1987). Transfer-sensitive inequality measures. *Review of Economic Studies 54*, 485–498.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives 13*(2), 145–166.

Sonne-Schmidt, C., F. Tarp, and L. P. Østerdal (2016). Ordinal bivariate inequality: Concepts and application to child deprivation in Mozambique. *Review of Income and Wealth 62*(3), 559–573.

Stiglitz, J. E., A. K. Sen, and J. P. Fitoussi (2009). Report by the commission on the measurement of economic performance and social progress. http://www.stiglitz-sen-fitoussi.fr/en/index.htm, Commission on the Measurement of Economic Performance and Social Progress.

Stowasser, T., F. Heiss, D. McFadden, and J. Winter (2014). Understanding the ses gradient in health among the elderly: The role of childhood circumstances. Technical Report Discussion Papers in Economics 20847, University of Munich, Department of Economics.

Strauss, J. and D. Thomas (1998). Health, nutrition, and economic development. *Journal of Economic Literature 36*(2), 766–817.

Trannoy, A. (2006). Multidimensional egalitarianism and the dominance approach: a lost paradise? In F. Farina and E. Savaglio (Eds.), *Inequality and Economic Integration.* London: Routledge.

Tsui, K.-Y. (1995). Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. *Journal of Economic Theory 67*, 251–265.

Tsui, K.-Y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Social Choice and Welfare 16*.

Zheng, B. (2008). Measuring inequality with ordinal data: a note. *Research on Economic Inequality 16*, 177–188.

Zheng, B. (2011). A new approach to measure socioeconomic inequality in health. *Journal Of Economic Inequality 9*, 555–577.

Appendix

Theorem 1 *Utilitarian unanimity of non-decreasing functions* $\iff$ *Generalised majorization* $\iff$ *LTSM Majorization*

Proof. We will start with the second equivalence.

$\Rightarrow$

Let $f \precsim_{LTSM} g$, then $f = Lg$ for some $LTSM$ L, then $\int f = \int Lg$, which implies $\int f \leqslant_{vec} \int g$[5] and we have $f \precsim g$.

$\Leftarrow$

For the converse we assume $f \precsim g$ and derive an $LTSM$ L such that $f = Lg$. Let us assume $f \neq g$, otherwise L would be identity matrix. We will now define a sequence of transfers that transform a distribution g into a distribution f. Let us find lowest p, q such that $f_p \neq g_p$ and $f_q \neq g_q$ (so we have $f_1 = g_1, ..., f_{p-1} = g_{p-1}$). There has to exist at least two such elements because if their sum is equal, they cannot differ on one position only. Recall that we have $f \precsim g$ so $f_p \leqslant g_p$ and $\frac{f_p}{g_p} \in (0, 1]$. We define $\varepsilon = (1 - \frac{f_p}{g_p}) \in [0, 1)$. We have $g^* = L_{pq}(\varepsilon)g = (g_1, g_2, ..., g_{p-1}, f_p, g_{p+1}, ..., g_{q-1}, (g_q - g_p + f_p), g_{q+1}, ..., g_n)$. Since $f_p \leqslant g_p$ we have also $f \leqslant_{vec} g^* \leqslant_{vec} g$ (so multiplication by $L_{pq}(\varepsilon)$ preserves majorization) and g^* agrees with f on one more position than g. We can now apply our algorithm as an induction step and by at most n exchanges we will obtain f ($f = g^{*(n \text{times})}$). Each such transfer defines elementary exchange matrix (Definition 8). The multiplication of two exchange matrices generates $LTSM$ matrix which is a consequence of Lemma 1.

Lemma 1 *LTSM matrices form a semigroup.*

Proof. It is clear that identity matrix belongs to $LTSM$. Now we only need to check that product of two $LTSM$ matrices is an $LTSM$ matrix. Let L and K be $LTSM$ matrices.

We have $(LK)_{ij} = l_{i1}k_{1j} + l_{i2}k_{2j} + ... + l_{in}k_{nj}$. Now remember that $l_{im} = 0$ for $m > i$ and $k_{mj} = 0$ for $m < j$, so if we want to obtain nonzero entries we need to have $j \leqslant m \leqslant i$. So $(LK)_{ij} = 0$ for $j > i$ (LK is lower triangular) and $(LK)_{ij} = l_{ij}k_{jj} + l_{i(j+1)}k_{(j+1)j} + l_{i(i-1)}k_{(i-1)j} + l_{ii}k_{ij}$ for $j \leqslant i$.

[5] $f = Lg = (l_{11}g_1, l_{21}g_1 + l_{22}g_2, ..., l_{n1}g_1 + l_{n2}g_2 + ... + l_{nn}g_n)$, $\int f = (l_{11}g_1, [l_{11} + l_{21}]g_1 + l_{22}g_2, ..., [l_{11} + l_{21} + .. + l_{n1}]g_1 + [l_{22} + l_{32} + ... + l_{n2}]g_2 + ... + l_{nn}g_n)$ and we clearly have $\int f \leqslant_{vec} \int g$ because each $[l_{1m} + l_{2m} + .. + l_{km}] \leqslant 1$, because L is column stochastic. Additionally we obtain that $\Sigma_{i=1}^n f_i = [l_{11} + l_{21} + .. + l_{n1}]g_1 + [l_{22} + l_{32} + ... + l_{n2}]g_2 + ... + l_{nn}g_n = \Sigma_{i=1}^n g_i$

We check that LK is column stochastic.[6]

$$\Sigma_{k=1}^{n}(LK)_{kj} = \Sigma_{t=j}^{n}(LK)_{tj} = \Sigma_{t=j}^{n}\Sigma_{m=j}^{t}l_{tm}k_{mj} =$$

$$\Sigma_{m=j}^{n}k_{mj}(\Sigma_{t=m}^{n}l_{tm}) = \Sigma_{m=j}^{n}k_{mj} = 1$$

since both $\Sigma_{m=j}^{n}k_{mj}$ and $\Sigma_{t=m}^{n}l_{tm}$ are equal to 1 due to L and K being an LTSM. We can now say that $f = Ly = L_{(n-1)n}(\varepsilon_{n-1})...L_{23}(\varepsilon_2)L_{12}(\varepsilon_1)g$ which finishes the proof.

To show first equivalence it is enough to write

$$W(u,f) = \Sigma_{i=1}^{n}u(i)f_i = \Sigma_{i=2}^{n}u(i)(F_i - F_{i-1}) + u(1)F_1 = \Sigma_{i=1}^{n-1}(u(i) - u(i+1))F_i + u(n)F_n$$

It is clear that $0 > W(u,f) - W(u,g) = \Sigma_{i=1}^{n-1}(u(i) - u(i+1))(F_i - G_i)$ for all $u \in \mathcal{U}_{nd}$ if and only if $F_i \leq G_i$ for all $i < n$

□

Therefore, if generalised majorization holds, some $LTSM$ matrix exists, namely, such that it is defined by the multiplication of the sequence of exchange matrices, which finishes the proof. □

Lemma 3 *$LTSM^2$ matrices form a semigroup.*

Proof. It follows directly from $\partial = \int^{-1}$ and the fact that $LTSM$ matrices form a semigroup. □

Theorem 2 *Utilitarian unanimity of concave functions $\iff$ Generalised majorization of 2nd order $\iff$ $LTSM^2$ Majorization*

Proof. We start with the second equivalence. Let us take $F = \int f$ and $G = \int g$. We obtain $\int F \leqslant_{vec} \int G$. By Theorem 2 we know that since $F \precsim G$ there exist LTSM L such that $F = LG$. Thus $f = \partial L \int g$.

Now we proceed to the first equivalence. Let us write

$$W(u,f) = \Sigma_{i=1}^{n}u(i)f_i = u(1)F_1 + \Sigma_{i=2}^{n}u(i)(F_i - F_{i-1}) = \Sigma_{i=1}^{n-1}(u(i) - u(i+1))F_i + u(n)F_n =$$

$$= \Sigma_{i=1}^{n-2}(u(i+2) - 2u(i+1) + u(i))(\Sigma_{j=1}^{i}F_j) + (u(n-1) - u(n))(\Sigma_{j=1}^{n-1}F_j) + u(n)F_n$$

[6] We regroup terms by k.

Since $\Sigma_{j=1}^{n-1} F_j = \Sigma_{j=1}^{n-1} G_j$ we write

$$W(u,f) - W(u,g) = \Sigma_{i=1}^{n-2}(u(i+2) - 2u(i+1) + u(i))(\Sigma_{j=1}^{i} F_j - G_j)$$

and it is easy to see that if $W(u,f) - W(u,g) \leq 0$ for all utility functions $u \in \mathcal{U}_{ndc}$, then $\Sigma_{j=1}^{i} F_j - G_j \geq 0$ for all i. □

Lemma 4 *Each $L^2 \in LTSM^2$ can be decomposed into at most $n-1$ matrices $L_{pq}^2(\varepsilon)$ exchange matrices (Definition 12).*

Proof. Let us take

$$\begin{aligned} G^2 \ni L^2 = \partial L \int &= \partial L_{(n-1)n}(\varepsilon_{n-1})...L_{23}(\varepsilon_2)L_{12}(\varepsilon_1) \int = \\ &= \partial L_{(n-1)n}(\varepsilon_{n-1}) \int \partial \ldots \int \partial L_{23}(\varepsilon_2) \int \partial L_{12}(\varepsilon_1) \int = \\ &= L_{(n-1)n}^2(\varepsilon_{n-1}) \ldots L_{23}^2(\varepsilon_2) L_{12}^2(\varepsilon_1) \end{aligned}$$

□

Lemma 4 *Each $L^2 \in LTSM^2$ can be decomposed into at most $n-1$ matrices $L_{pq}^2(\varepsilon)$ exchange matrices (Definition 12).*

Proof. Let $D_{pq}(\varepsilon) = L_{pq}(\varepsilon) - Id$, i. e. $(D_{pq})_{pp} = -\varepsilon$, $(D_{pq})_{qp} = \varepsilon$ and $(D_{pq})_{ij} = 0$ elsewhere. $L_{pq}^2(\varepsilon) = \partial L_{pq}(\varepsilon) \int = \partial(Id + D_{pq}(\varepsilon)) \int = \partial Id \int + \partial D_{pq}(\varepsilon) \int = Id + \partial D_{pq}(\varepsilon) \int$ and it immediately follows that L_{pq}^2 is of the above form. □

Theorem 3 *Utilitarian unanimity of non-decreasing concave functions $\iff$ Generalised nondecreasing majorization of 2nd order $\iff$ $sLTSM^2$ Majorization*

Proof. Let us start with the second equivalence. The right implication follows from the fact that $L^2 = \partial L \int$ and proof of Theorem 1. To prove left implication, we assume that $f \precsim_{nd}^2 g$ and derive $sLTSM^2$ matrix L^2. We assume that $f \neq g$, otherwise $L^2 = Id$. Let us find the lowest p such that $F_p \neq F_p$ and if possible the lowest q such that $p < q$ and $F_q \neq G_q$. We have

$f_1 = g_1, \dots, f_{p-1} = g_{p-1}$ by definition of p and since $f \precsim^2 g$ we have also $f_p < g_p$. If such q exists and $g_{q+1} < g_p - f_p$ we introduce $\varepsilon^s = \frac{g_p - f_p - g_{q+1}}{G_p}$. We have $g^* = L_p^{s\,2}(\varepsilon^s) = (g_1, g_2, \dots, g_{p-1}, f_p + g_{q+1}, g_{p+1} + g_p - f_p - g_{q+1}, g_{p+2} \dots)$ and further $G^* = (G_1, G_2, \dots, G_{p-1}, F_p + g_{q+1}, G_{p+1}, G_{p+2} \dots)$ and since $f_p \leq f_p + g_{q+1} \leq g_p$ we have $f \precsim^2 g^* \precsim^2 g$, and $g_p^* - f_p = g_{q+1}^*$. Now we can assume that $g_{q+1} \geq g_p - f_p$. We will introduce $\varepsilon = 1 - \frac{F_p}{G_p}$ and first consider the case when $q = p+1$. We have $g^* = L_{p(p+1)}^2(\varepsilon)g = (g_1, g_2, \dots, g_{p-1}, f_p, g_{p+1} + 2G_p - 2F_p, g_{p+2} + F_p - G_p, g_{p+3} \dots)$, and further $G^* = (G_1, G_2, \dots, G_{p-1}, F_p, G_{p+1} + G_p - F_p, G_{p+2} \dots)$, Let $\mathcal{F} = \int F$, $\mathcal{G}^* = \int G^*$ and $\mathcal{G}^* = \int G^*$. We have $\mathcal{G}* = (\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{F}_p, \mathcal{G}_{p+1}, \dots)$ and it is obvious that $f \precsim^2 g^* \precsim^2 g$. Now let us consider $q > p+1$. We have $g^* = L_{pq}^2 = (\varepsilon)g = (g_1, g_2, \dots, g_{p-1}, f_p, g_{p+1} + G_p - F_p, g_{p+2}, \dots, g_{q-1}, g_q + G_p - F_p, g_{q+1} - G_p + F_p, g_{q+2} \dots)$, and further $G^* = (G_1, G_2, \dots, G_{p-1}, F_p, G_{p+1}, \dots, G_{q-1}, G_q + G_p - F_p, G_{q+1}, \dots)$, We obtain further $\mathcal{G}* = (\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{F}_p, \mathcal{F}_{p+1}, \dots, \mathcal{F}_{q-1}, \mathcal{G}_q, \mathcal{G}_{q+1}, \dots)$ and it is obvious that $f \precsim^2 g^* \precsim^2 g$. Now let us consider the case when such q does not exist. We have $g^* = L_p^{s\,2}(\varepsilon) = (g_1, g_2, \dots, g_{p-1}, f_p, g_{p+1} + G_p - F_p, g_{p+2}, \dots)$ and further $G^* = (G_1, G_2, \dots, G_{p-1}, F_p, G_{p+1}, G_{p+2} \dots)$ and it is again obvious that $f \precsim^2 g^* \precsim^2 g$. This way we constructed a distribution g^* such that $f \precsim^2 g^* \precsim^2 g$ and g^* agrees with f on one more category. We finish our proof by induction.

To prove first equivalence it is enough to show that if $f = L_{pq}^2(\varepsilon)g$, then $W(u,f) - W(u,g) = \varepsilon G_p(-u(p) + u(p+1) + u(q) - u(q+1))$ and if $f = L_p^{s\,2} g$ then $W(u,f) - W(u,g) = \varepsilon G_p(-u(p) + u(p+1))$. To show converse implication we take $u(1) = \dots = u(p) \leq u(p+1) \leq u(p+2) = \dots = u(n)$ such that $u(p+2) - 2u(p) + u(p) \leq 0$ and show by induction that $f \precsim_{nd}^2 g$. □

Lemma 5 *For a given partial ordering $\widetilde{LTSM}$ consists of $LTSM$ matrices such that if $\zeta_{ij} = 0$ then $l_{ij} = 0$.*

Proof. Let us take two matrices $L_1, L_2 \in \widetilde{LTSM}$ such that $(L_1)_{ts} > 0$, $(L_2)_{ts} > 0$ if $\zeta_{ts} = 1$ and $(L_1)_{ts} = (L_2)_{ts} = 0$ if $\zeta_{ts} = 0$. Let $L = L_1 L_2$. To prove our theorem, it is enough to show that for any t, s if $\zeta_{ts} = 0$ then $l_{ts} = 0$. Let us choose such $t > s$ (if $t \leq s$ implication follows from L being lower triangular matrix). Let us calculate $l_{ts} = \Sigma_{k=1}^{nm}(L_1)_{tk}(L_2)_{ks} = \Sigma_{k=s+1}^{t-1}(L_1)_{tk}(L_2)_{ks}$, with second equality following from L_1, L_2 being lower triangular matrices and $(L_1)_{ts} = (L_2)_{ts} = 0$. Now we should note that for $s \leq k \leq t$ we have either $(L_1)_{tk} = 0$ or $(L_2)_{ks} = 0$ ($(L_1)_{tk}(L_2)_{ks} = 0$) and to check this we will assume that $(L_1)_{tk}(L_2)_{ks} > 0$. From our previous assumptions we therefore know that $\zeta_{tk} = \zeta_{ks} = 1$ and hence $(i_k, j_k) \precsim (i_t, j_t)$, $(i_s, j_s) \precsim (i_k, j_k)$. Now, by the transitivity we have $(i_s, j_s) \precsim (i_t, j_t)$ and so $\zeta_{ts} = 1$ which contradicts our assumption that $\zeta_{ts} = 0$. Since each component of the sum is equal to 0 we have $l_{ts} = 0$ which finishes our proof. □

Theorem 4 *$\widetilde{LTSM}$ Majorization ⇒ Bidimensional generalised orthant majorization ⟺ Utilitarian unanimity of non-decreasing and submodular functions*

Proof. The proof of the right implication is identical to the proof of the right implication in Theorem 1. To prove the equivalence we write

$$\begin{aligned} W(u,f) = &\Sigma_{i=1}^{n}\Sigma_{j=1}^{m} u(i,j) f_{ij} = \\ &= \Sigma_{i=1}^{n-1}\Sigma_{j=1}^{m-1}(u(i+1,j+1) - u(i,j+1) - u(i+1,j) + u(i,j))F_{ij} + \\ &+ \Sigma_{i=1}^{n-1}(u(i,m) - u(i+1,m))F_i^1 + \Sigma_{j=1}^{m-1}(u(n,j) - u(n,j+1))F_i^1 + u(n,m) \end{aligned} \tag{4}$$

From this equation it is easy to see that $W(u,f) - W(u,g) \leq 0$ for all utility functions $u \in \mathcal{U}_{ndsub}$[7] if and only if $\Sigma_{k=1}^{i}\Sigma_{l=1}^{j} F_{kl} - G_{kl} \geq 0$ for all i, j. □

Lemma 6 *If distributions f and g have the same marginal distributions and $\tilde{\int} f \leq_{vec} \tilde{\int} g$ then $f = g$.*

Proof. We will assume that they are not equal and prove a contradiction. Let us find lowest (i, j) (there is no (k, l) such that $(k, l) \precsim (i, j)$) such that $f_{ij} < g_{ij}$ - let it be (i_1, j_1). Then we find position with highest i, j such that $f_{ij_1} > g_{ij_1}$ or $f_{i_1 j} > g_{i_1 j}$, let them be i_2, j_2. Now we define following comprehensive sets $K_1 = \{(k,l) \in \mathbb{N}^2 | k \leq i_2,\ l < j_1\}$, $K_2 = \{(i,j) \in \mathbb{N}^2 | i < i_1,\ j \leq j_2\}$, and denote $\Sigma_{(i,j)\in K} f_{ij} - g_{ij}$ by $S(K)$. We have $S(K_1) = \Sigma_{i=i_1}^{i_2} f_{ij_1} - g_{ij_1} = \Sigma_{i=1}^{i_2} f_{ij_1} - g_{ij_1} < 0$, and from this and same marginals we get $\Sigma_{i=i_2+1}^{n} f_{ij_1} - g_{ij_1} > 0$ which means $i_2 = n$ (or else it is contradiction to the fact that i_2 is highest index for which $f_{ij_1} > g_{ij_1}$). Analogously $j_2 = m$ and we define $K_3 = K_1 \cup K_2$ and calculate $S(K_3) = \Sigma_{i=i_1}^{n} f_{ij_1} - g_{ij_1} + \Sigma_{j=j_1+1}^{m} f_{i_1 j} - g_{i_1 j} = f_{j_1}^2 - g_{j_1}^2 + (f_{j_1}^2 - f_{i_1 j_1}) - (g_{j_1}^2 - g_{i_1 j_1}) = g_{i_1 j_1} - f_{i_1 j_1} > 0$ which gives us contradiction. □

[7] If $u(i+1,j+1) - u(i,j+1) - u(i+1,j) + u(i,j) < 0$ for all $i < n$, $j < m$ and $u(i,m) - u(i+1,m) < 0$ for all $i < n$ then for chosen i we have $u(i,j) - u(i+1,j) < 0$, and similarly for the other variable

Theorem 5 *Utilitarian unanimity of non-decreasing functions* $\iff$ *Bidimensional generalised majorization* $\iff$ $\widetilde{LTSM}$ *Majorization*

Proof. Bidimensional generalised majorization $\Longrightarrow$ $\widetilde{LTSM}$ *Majorization*

We will construct an algorithm to make such a sequence of elementary transformations. First let us calculate $d_{ij} = f_{ij} - g_{ij}$, our goal is to make $d_{ij} = 0$ for all (i, j) by using elementary transformations - we will start from calculating $d^1_i = \Sigma^m_{j=1} d_{ij} = f^1_i - g^1_i$, $d^2_j = \Sigma^n_{i=1} d_{ij} = f^2_j - g^2_j$. We will concentrate on the first marginal. We start from finding a lowest i such that d^1_i is negative and finding a highest j, such that $d_{i_1 j_1}$ is negative. Then we try to find lowest possible i bigger than i_1 such that there exist $j \geq j_1$ for which d_{ij} is positive, and from such j we choose the highest, let us call them i_2, j_2. Such a pair has to exist because sums of differences over sets complementary to comprehensive sets is of the opposite sign than in the case of comprehensive set. We will transfer at most $\min(d_{i_2 j_2}, -d_{i_1})$ from (i_2, j_2) to $(i_2 - 1, j_2)$, without worrying about what was before at $d_{(i_2-1)j_2}$. It is clear that our new distribution $\tilde{f}$ fulfills $\tilde{\int} f \leq_{vec} \tilde{\int} \tilde{f}$ - if (i_2, j_2) does not belong to comprehensive set K then sum has no less value for $\tilde{f}$ - they differ only for two pairs. If (i_2, j_2) belongs to K then so does $(i_2 - 1, j_2)$ and sums are equal. Our biggest concern is then if $\tilde{d}_{ij}$ still give us negative value while summing over comprehensive sets. Let us use notation of $S(K)$ from lemma 6. Let us assume that (i_2, j_2) does not belong to K but (i_1, j_2) does, otherwise we have $\tilde{S}(K) = S(K) < 0$. We have $\tilde{S}(K) = S(K) + \min(d_{i_2 j_2}, -d_{i+1})$. Let assume that $S(K) + d_{i_2 j_1} > 0$ for some comprehensive set K and that it is highest possible value - there is no set L such that $S(K) < S(L)$ and L fulfils conditions for (i_1, j_1) and (i_2, j_1). Obviously since $\Sigma^m_{j=k} d_{ij} \geq 0$ for $i < i_1$ (because $d^1_i = 0$) we have that K contains all (i, j) such that $i < i_1$ and since j_1 is highest possible such that $d_{i_1 j}$ is negative, we can say that K contains all pairs (i, j) such that $i \leq i_1$. Since K contains $(i_2 - 1, j_2)$, it also contains all of the pairs (i, j) such that $i_1 < i < i_2$, $j \leq j_2$. It does not contain any pair such that $j > j_2$. since from definition of i_2 they all are negative. Finally since $d_{i_2 j_2}$ is the highest positive with respect to j then above j_2 are only negative or neutral and so we can expand our set K by the pairs (i_2, j) such that $j < j_2$. But now we see that $S(K) + \min(d_{i_2 j_2}, -d_{i+1}) > 0$ is contradictory to $S(K) + d_{i_2 j_2} < 0$, that is the same as if we added (i_2, j_2) to set K, obtaining set $\tilde{K}$ that has $S(\tilde{K}) < 0$ which concludes our inductive step. Now we repeat this step, until $d^1_{i_1} = 0$, then proceed to the next point of the marginal distribution, reaching the same marginals in the finite number of steps. We use lemma 6 to say that $f_n = g$.

$\widetilde{LTSM}$ Majorization $\Longrightarrow$ Utilitarian unanimity of non-decreasing functions
Let $d^k(i,j) = f_k - f_{k-1}$, that is d^k is an elementary transformation along first ($k = 1$) or second ($k = 2$) marginal, making transfer from (i,j) to $(i+1,j)$ or from (i,j) to $(i,j+1)$. We have for $d^1(i,j)$

$$W(u,f_k)-W(u,f_{k-1}) = u(i,j)((f_k)_{ij}-(f_{k-1})_{ij})+u(i+1,j)((f_k)_{(i+1)j}-(f_{k-1})_{(i+1)j}) = \varepsilon(u(i,j)-u(i+1,j))$$

and similarly for $d^2(i,j)$

$$W(u,f_k) - W(u,f_{k-1}) = \varepsilon(u(i,j) - u(i,j+1))$$

and it is clear that all functions $u \in \mathcal{U}_{nd}$ fulfil inequality.

Utilitarian unanimity of non-decreasing functions $\Longrightarrow$ Bidimensional generalised majorization
Let $u(i,j) = \begin{cases} 0 & for(i,j) \in K \\ 1 & for(i,j) \notin K \end{cases}$, then $0 \leq= W(u,f)-W(u,g) = \Sigma_{(i,j)\notin K} f_{ij}-g_{ij} = \Sigma_{(i,j)\in K} g_{ij}-f_{ij}$ for each comprehensive set K. □

Theorem 6 *$\widetilde{LTSM}^2$ Majorization $\Leftrightarrow$ Bidimensional generalised majorization of 2nd order $\Leftrightarrow$ Utilitarian unanimity*

Proof. Let us rewrite theorem as

$$\widetilde{\int}\int x \leq_{vec} \widetilde{\int}\int y \iff \exists_{\tilde{L}^2 \in \widetilde{LTSM}^2} x = \tilde{L}^2 y$$

Now we substitute X for $\int x$ and Y for $\int y$ and from Theorem 4 we get

$$\widetilde{\int} X \leq_{vec} \widetilde{\int} Y \iff \exists_{\tilde{L} \in s\widetilde{LTSM}} X = \tilde{L} Y.$$

From the fact that for all $\tilde{L} \in s\widetilde{LTSM}$ there exists $\tilde{L}^2 \in \widetilde{LTSM}^2$ such that $\tilde{L}^2 = \partial \tilde{L} \int$ we obtain our result. □