Decomposing bivariate dominance for social welfare comparisons

by

Tina Gottschalk, Troels Martin Range, Peter Sudhölter and Lars Peter Østerdal

Discussion Papers on Business and Economics
No. 12/2015

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55, DK-5230 Odense M
Denmark

E-mail: lho@sam.sdu.dk / http://www.sdu.dk/ivoe

Decomposing bivariate dominance for social welfare comparisons*

Tina Gottschalk† Troels Martin Range‡ Peter Sudhölter§

Lars Peter Østerdal¶

Department of Business and Economics, and COHERE, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

June 17, 2015

JEL classification: C63, D63, I31

Keywords: social welfare, lower orthant dominance, first order dominance, algorithm

Abstract

The lower orthant dominance relation is frequently used for multidimensional social welfare comparisons. Recently it has been shown that bivariate dominance can be characterized in terms of elementary mass transfer operations. We provide an algorithm which explicitly decomposes the mass transfers into welfare differences and inequality differences.

1 Introduction

Dominance concepts are increasingly used for multidimensional comparisons of social welfare, inequality and poverty (Aaberge and Brandolini 2014).[1] They provide a way of making robust comparisons of the overall attainment of the groups. In particular, comparisons do not require assumptions about the relative importance of each of the given dimensions.

A simple and commonly used dominance concept for multidimensional social welfare comparisons is lower orthant dominance. The idea of using orthant dominance – and related (less restrictive) concepts – for social welfare comparisons was popularized by Atkinson and Bourguignon (1982), and it has been significantly developed and refined in several works, (see, e.g., Bourguignon and Chakravarty 2003, Duclos, Sahn and Younger 2006, Duclos, Sahn and Younger 2007, Gravel, Moyes and Tarroux 2009, Gravel and

*Support from the Danish Council for Independent Research | Social Sciences (Grant-id: DFF – 1327-00097) is acknowledged, and the third author acknowledges support from the Spanish Ministerio de Ciencia e Innovación under project ECO2012-33618, co-funded by the ERDF.

†Email: tigo@sam.sdu.dk

‡Email: tra@sam.sdu.dk

§Email: psu@sam.sdu.dk

¶Email: lpro@sam.sdu.dk

[1] Stochastic dominance is useful in many other fields than welfare economics. It is, for example, an important tool in decision theory (see, e.g., Levy 1992 or Müller and Stoyan 2002), finance (see, e.g., Sriboonchita, Dhompongsa, Wong, and Nguyen 2009), as well as in probability theory and statistics (see, e.g., Silvapulle and Sen 2011).

Mukhopadhyay 2010 and Muller and Trannoy 2011).[2] Suppose that, for each dimension, a wellbeing indicator can take a finite number of possible values. Considering multiple dimensions of wellbeing, the set of possible (multidimensional) outcomes is thus the product set of those possible indicator values for each dimension. We describe a population distribution by a probability mass function over the outcomes, i.e., by a function that assigns to each outcome the probability that a randomly selected individual obtains that outcome (or, put differently, it describes the *share* of all individuals in the population that obtains that outcome). For two probability mass functions (i.e., population distributions) f and g, f *(lower orthant) dominates* g if and only if the cumulative probability mass at f is smaller than or equal to that at g for every lower hyperrectangle. This is equivalent to the property that the expected utility (welfare) of f is at least as high as that of g for any non-decreasing utility (welfare) function with negative cross derivative (Atkinson and Bourguignon 1982).

The aforementioned characterization of lower orthant dominance is well-known and central to the literature. The latter definition has a foundation in expected utility theory (and social welfare maximization),[3] while the former, equivalent, definition has an operational meaning as it provides an easy way of checking dominance. However, until very recently, a characterization founded in elementary operations, i.e. conditions specifying exactly which changes in a distribution that are allowed to obtain another distribution which dominates it, has been missing.[4] Such a characterization was recently given by Meyer and Strulovici (2015) who, in particular for the bivariate case, showed that one probability mass function dominates another if and only if the dominated probability mass function can be obtained from the dominating probability mass function by diminishing probability mass transfers and correlation increasing switches.[5] Here, a diminishing probability mass transfer is simply a shift of mass from a better to a worse outcome, hence such a transfer implies an unambiguous welfare reduction.[6] A correlation increasing switch consists of two simultaneous transfers that move mass from intermediate outcomes to more extreme outcomes without changing the marginal distributions. Among others, Atkinson and Bourguignon (1982), Tsui (1999), Decancq (2012) and Sonne-Schmidt, Tarp, and Østerdal (2015) argue that correlation increasing switches are basic operations that (unambiguously) increase inequality.

In this paper we show how lower orthant dominance decomposes into welfare reductions and increases in inequality for the two-dimensional case. More precisely, we provide a *constructive* proof of Meyer and Strulovici's characterization, which, most importantly, yields an algorithm that returns a set of diminishing transfers and correlation increasing switches whenever a dominance relationship exist. This means that we can meaningfully disentangle the differences between the two distributions, in term of a component increasing inequality (the correlation increasing switches) and a component decreasing welfare (the diminishing transfers). The decomposition technique also enables us to compare lower orthant dominance to more restrictive stochastic dominance concepts – in particular to the usual stochastic order which allows only diminishing transfers.

[2]Note that lower orthant dominance has sometimes been referred to as "first order dominance", particularly in the welfare economics literature. In order not to risk confusion with the usual stochastic order – the natural dominance concept for social welfare comparisons with ordinal data (see, e.g., Arndt et al. 2012, Østerdal 2010, and Range and Østerdal 2013) – we use the term lower orthant dominance as conventional in the probability theory literature (e.g. Shaked and Shanthikumar 2007).

[3]The negative cross derivative can be interpreted as an assumption about complementarity of dimensions.

[4]For example, Moyes (2012) points out in his Footnote 13 that such characterization is missing, even though there are results in the literature that are making steps in this direction.

[5]Meyer and Strulovici (2015) is based on a 2010 draft entitled "The Supermodular Stochastic Ordering". Müller (2013) provides a similar characterization by elegant use of duality theory.

[6]Indeed, the usual stochastic order is completely characterized by such transfers, as shown by, e.g., Strassen (1965) and Kamae, Krengel, and O'Brien (1977).

2 Notation

Let $n, m \in \mathbb{N}$. For $x, y \in \mathbb{R}^m$, $x \leqslant (\geqslant)\ y$ denotes $x_i \leqslant (\geqslant)\ y_i$ for all $i = 1, \ldots, m$, $x < (>)\ y$ means $x \leqslant (\geqslant)\ y$ and $x \neq y$, and $x \ll (\gg)\ y$ if $x_i < (>)\ y_i$ for all $i = 1, \ldots, m$. Similarly, for two functions $f, g : D \to \mathbb{R}$, $f \geqslant (\leqslant)\ g$ if $f(x) \geqslant (\leqslant)\ g(x)$ for all $x \in D$, and $f > (<)\ g$ if $f \geqslant (\leqslant)\ g$ and $f \neq g$.

Denote $X(n, m) = X = \{x \in \mathbb{N}^2 \mid x \leqslant (n, m)\}$ and $\mathcal{F}(n, m) = \mathcal{F} = \{f : X \to \mathbb{R}_+\}$. For $\emptyset \neq Y \subseteq X$ let $\max Y = y \in X$ be defined by $y_i = \max\{x_i \mid x \in Y\}$ for $i = 1, 2$, and let $\min Y$ be defined analogously. Moreover, for $x \in X$, $\downarrow x = \{y \in X \mid y \leqslant x\}$. For $f, g \in \mathcal{F}$ we say that g results from f

- by a *diminishing (bilateral) transfer* if there exist $x, y \in X$ such that $x < y$, $g(x) - f(x) = f(y) - g(y) > 0$, and $g(z) = f(z)$ for all $z \in X \setminus \{x, y\}$ (the underlying transfer is a transfer *from* y *to* x of size $\varepsilon = g(x) - f(x)$) and we use the notation $g = f_\varepsilon^{x \leftarrow y}$;
- by a *correlation increasing switch* if there exist $x, y \in X$ such that $f(x) - g(x) = f(y) - g(y) = g(v) - f(v) = g(w) - f(w) > 0$, and $f(z) = g(z)$ for all $z \in X \setminus \{x, y, v, w\}$, where $v = \min(\{x, y\})$ and $w = \max(\{x, y\})$ (note that in this case x and y are incomparable, i.e., $x \not\leqslant y \not\leqslant x$, and that the underlying switch transfers $\varepsilon = f(x) - g(x)$ from each x and y to each v and w) and we use the notation $g = f_\varepsilon^{x \leftrightarrows y}$.

Diminishing transfer is illustrated in Figure 2.1 and correlation increasing switch is illustrated in Figure 2.2.

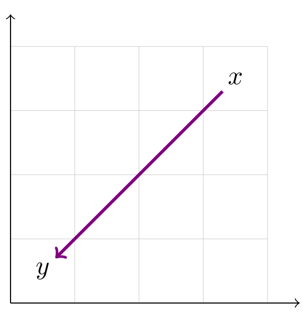

Figure 2.1: Diminishing transfer

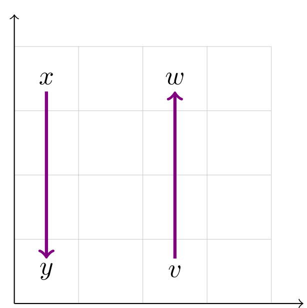

Figure 2.2: Correlation increasing switch

Throughout the following notation is employed. For any $f \in \mathcal{F}(n, m)$ and $x \in X(n, m)$ denote $\widetilde{f}(x) = \sum_{y \in \downarrow x} f(y)$, i.e., the cumulative distribution where we for each entry sum all probability mass below and left of the entry including the row and column in line with the entry, and $\widetilde{f}(0, i) = \widetilde{f}(j, 0) = f(0, 0) = 0$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. Furthermore we will denote the marginal distribution as the partial distribution of each dimension. We can find the marginal distributions by looking respectively at $\widetilde{f}(n, i)$ and $\widetilde{f}(j, m)$.

Definition 2.1 *Let $n, m \in \mathbb{N}$ and $f, g \in \mathcal{F}(n, m)$. We say that f* weakly dominates *g, written $f \succeq g$, if there exist $k \in \mathbb{N}$ and $f_1, \ldots, f_k \in \mathcal{F}(n, m)$ such that $f = f_1$, $g = f_k$, and, for all $\ell \in \{2, \ldots, k\}$, f_ℓ results from $f_{\ell-1}$ by a diminishing transfer or by a correlation increasing switch. Moreover, we say that f* lower orthant dominates *g, written $f \succeq_{LO} g$, if $\widetilde{g}(n, m) = \widetilde{f}(n, m)$ and $\widetilde{g} \geqslant \widetilde{f}$. Finally, let $\succ$ and $\succ_{LO}$ denote the respective strict relations (requiring in addition $k > 1$ and $\widetilde{g} > \widetilde{f}$, respectively).*

3 Results and algorithm

We are now able to formulate the characterization by Meyer and Strulovici (2015).

Theorem 3.1 *Let $n, m \in \mathbb{N}$ and $f, g \in \mathcal{F}(n,m)$. $f \succeq g$ if and only if $f \succeq_{LO} g$.*

Our proof is constructive and will proceed in several steps. We will start by showing Proposition 3.2, the straightforward direction of the equivalence. We just consider what happens when we make a diminishing bilateral transfer or a correlation increasing switch. A diminishing probability transfer only moves mass in direction towards the origin, i.e., we only increase $\widetilde{f}$-values. Furthermore, a correlation increasing switch might decrease some $\widetilde{f}$-values, but because it is a correlation increasing switch, it will simultaneously increase the $\tilde{f}$-values somewhere else.

Proposition 3.2 *Let $n, m \in \mathbb{N}$ and $f, g \in \mathcal{F}(n,m)$. If $f \succeq g$, then $f \succeq_{LO} g$.*

Proof: Let $f, g \in X(n,m)$. If g results from f by a diminishing bilateral transfer from y to x or by a correlation increasing switch from x to $\min\{x,y\}$ and y to $\max\{x,y\}$, then $\widetilde{g}(n,m) = \widetilde{f}(n,m)$ and it is straightforward to show that $\widetilde{g} \geq \widetilde{f}$. Therefore, Proposition 3.2 follows by induction on k, the number of functions $f_1, \ldots, f_k \in \mathcal{F}$ such that $f_1 = f$, $f_k = g$, and f_ℓ results from $f_{\ell-1}$ by a diminishing transfer or a correlation increasing switch for each $\ell \in \{2, \ldots, k\}$. **q.e.d.**

Proposition 3.3 *Let $n, m \in \mathbb{N}$ and $f, g \in \mathcal{F}(n,m)$. If $f \succeq_{LO} g$, then $f \succeq g$.*

Our proof of Proposition 3.3 is constructive and yields an algorithm. First we construct a finite sequence of diminishing transfers creating a function $h, f \succeq h \succeq_{LO} g$, that has the same marginal distributions as g. Indeed, applying Lemma 3.4 successively to the rows $i_0 = m, \ldots, 1$ not only leads to a function h that has the same marginal distributions as g, but also assigns at least the same mass as g to each upper rectangle of the form $\{(j', i') \in X \mid j' \leqslant j, i' \geqslant i\}$ where $(j,i) \in X$ (see Corollary 3.5).

As every further diminishing transfer increases at least one marginal distribution, it results in a function that does no longer lower orthant dominate g. If still $h \succ_{LO} g$, then we construct in a second step a finite sequence of correlation increasing switches that finally leads to g.

Lemma 3.4 *Let $i_0 \in \{1, \ldots, m\}$ and $f \succeq_{LO} g$ such that*

$$\begin{aligned} \widetilde{f}(j,m) - \widetilde{f}(j,i) &\geqslant \widetilde{g}(j,m) - \widetilde{g}(j,i) \ \forall j \in \{1, \ldots, n-1\}, i \in \{i_0, \ldots, m\} \text{ and} && (1) \\ \widetilde{f}(n,i) &= \widetilde{g}(n,i) \ \forall i \in \{i_0, \ldots, m\}. && (2) \end{aligned}$$

Then there exists $h \in \mathcal{F}$ that arises from f by finitely many diminishing transfers such that $h \succeq_{LO} g$ and

$$\begin{aligned} \widetilde{h}(j,m) - \widetilde{h}(j,i) &\geqslant \widetilde{g}(j,m) - \widetilde{g}(j,i) \ \forall j \in \{1, \ldots, n-1\}, i \in \{i_0 - 1, \ldots, m\} \text{ and} && (3) \\ \widetilde{h}(n,i) &= \widetilde{g}(n,i) \ \forall i \in \{i_0 - 1, \ldots, m\}. && (4) \end{aligned}$$

Lemma 3.4 reveals the construction of the maximal sequence of diminishing transfers. We first illustrate and explain its proof. The proof is divided into two steps.

In the first step consider the set of all $j \in \{1, \ldots, n\}$ where (3) is not fulfilled and denote its largest index by j_0. In the 4×4 example of Fig. 3.1, $i_0 = 3, j_0 = 2$, and in the red area f has less probability mass than

g. However, f has the same amount or more probability mass than g in the blue area. This means that the difference between the probability masses of f and g in the grey area must be larger than or equal to the difference between the probability masses of f and g in the red area. Therefore this amount from the grey area can be moved to the red area by diminishing bilateral transfers. Now we have removed j_0 from the set of all $j \in \{1, \dots, n\}$ where (3) is not fulfilled, and we proceed recursively.

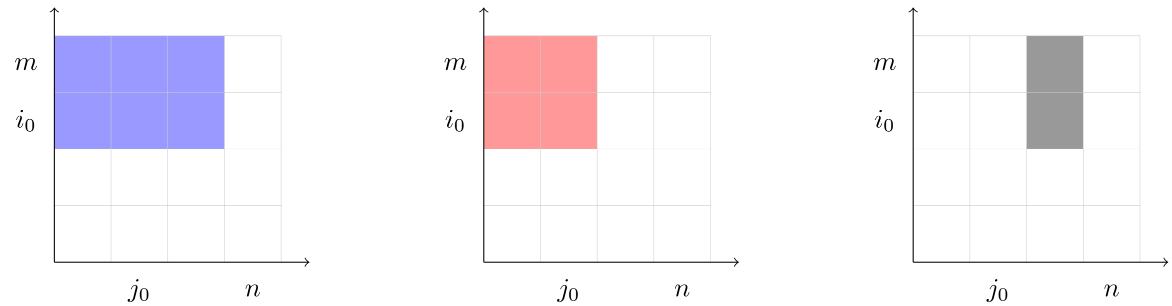

Figure 3.3: Illustration of *Step 1* in the proof of Lemma 3.4.

So far we have only employed diminishing transfers from $(j_0 + 1, i)$ to (j_0, i) for some $i \in \{i_0, \dots, m\}$. In the second step we have to make sure that $\widetilde{f}(n, i_0 - 1) = \widetilde{g}(n, i_0 - 1)$, i.e. the green area in Figure 3.4 must be equal in f and g. To ensure this we employ diminishing transfers from (j, i_0) to $(j, i_0 - 1)$ for some $j \in \{1, \dots, n\}$.

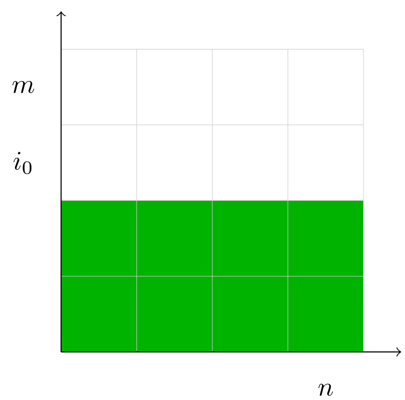

Figure 3.4: Illustration of *Step 2* in the proof of Lemma 3.4.

Proof of Lemma 3.4: Due to lower orthant dominance we have $\widetilde{g}(n, i_0 - 1) \geqslant \widetilde{f}(n, i_0 - 1)$ and $\widetilde{g}(n, m) = \widetilde{f}(n, m)$,

$$\widetilde{f}(n, m) - \widetilde{f}(n, i_0 - 1) \geqslant \widetilde{g}(n, m) - \widetilde{g}(n, i_0 - 1). \tag{5}$$

Step 1: We show that there exists h that arises from f by finitely many diminishing transfers such that $h \succeq_{LO} g$ and

$$\widetilde{h}(j, m) - \widetilde{h}(j, i) \geqslant \widetilde{g}(j, m) - \widetilde{g}(j, i) \tag{6}$$

for all $j \in \{1, \dots, n\}$ and all $i \in \{i_0 - 1, \dots, m\}$.

Let

$$K(f) = K = \{j \in \{1, \dots, n\} \mid \widetilde{f}(j, m) - \widetilde{f}(j, i_0 - 1) < \widetilde{g}(j, m) - \widetilde{g}(j, i_0 - 1)\}$$

and $k = |K|$. We proceed by induction on k. If $k = 0$, then $h = f$ fulfils (3). Assume that our statement is correct whenever $k < \ell$ for some $\ell \in \mathbb{N}$. Now, if $k = \ell$, we proceed as follows. Let

$j_0 = \max K$. By (5), $j_0 < n$. Moreover, because j_0 is maximal, $\alpha := \widetilde{f}(j_0+1, m) - \widetilde{f}(j_0+1, i_0-1) \geqslant \widetilde{g}(j_0+1, m) - \widetilde{g}(j_0+1, i_0-1) := \beta$, we conclude that

$$\begin{aligned} &\Big(\alpha - \big(\widetilde{f}(j_0, m) - \widetilde{f}(j_0, i_0-1)\big)\Big) - \Big(\beta - \big(\widetilde{g}(j_0, m) - \widetilde{g}(j_0, i_0-1)\big)\Big) \\ \geqslant\ &\big(\widetilde{g}(j_0, m) - \widetilde{g}(j_0, i_0-1)\big) - \big(\widetilde{f}(j_0, m) - \widetilde{f}(j_0, i_0-1)\big), \end{aligned}$$

i.e.,

$$\sum_{i=i_0}^{m} \Big(f(j_0+1, i) - g(j_0+1, i)\Big) \geqslant \big(\widetilde{g}(j_0, m) - \widetilde{g}(j_0, i_0-1)\big) - \big(\widetilde{f}(j_0, m) - \widetilde{f}(j_0, i_0-1)\big) =: \varepsilon > 0. \quad (7)$$

Let $A = \{i \in \{i_0, \ldots, m\} \mid f(j_0+1, i) > g(j_0+1, i)\}$. There exist $\varepsilon_i \geqslant 0, i \in A$, such that $\sum_{i \in A} \varepsilon_i = \varepsilon$ and $f(j_0+1, i) - \varepsilon_i \geqslant g(j_0+1, i)$ for all $i \in A$. Let h arise from f by transferring ε_i from (j_0+1, i) to (j_0, i) for all $i \in A$ (i.e., h arises from f by a sequence of $|A|$ diminishing transfers). As $\widetilde{g}(j_0, i_0-1) \geqslant \widetilde{f}(j_0, i_0-1)$, (1) applied to $j = j_0$ yields $\widetilde{g}(j_0, i) \geqslant \widetilde{f}(j_0, i) + \varepsilon$ for all $i = i_0, \ldots, m$. As $\widetilde{h}(j_0, i) - \widetilde{f}(j_0, i) \leqslant \varepsilon$ for $i \in A$ and $\widetilde{h}(j, i) = \widetilde{f}(j, i)$ for all other pairs (j, i), we conclude that $h \succeq_{LO} g$. Moreover, $K(h) \subseteq K(f) \setminus \{j_0\}$ by construction. Finally, as only transfers from the right hand to the left hand side are employed, (6) is still satisfied for all $i \in \{i_0, \ldots, m\}$ by (1). Hence, by the inductive hypothesis, this step is complete.
Step 2: We now finish the proof. By Step 1 we may assume that $\widetilde{f}(j, m) - \widetilde{f}(j, i_0-1) \geqslant \widetilde{g}(j, m) - \widetilde{g}(j, i_0-1)$ for all $j = 1, \ldots, n$. Let $\widetilde{g}(n, i_0-1) - \widetilde{f}(n, i_0-1) = \rho(f) = \rho \geqslant 0$. If $\rho = 0$, we may choose $h = f$. Hence we assume that $\rho > 0$. Let

$$L(f) = L = \{j \in \{1, \ldots, n\} \,|\, f(j, i_0) > g(j, i_0)\},$$

and $\ell = |L|$. As $\rho = \sum_{t=1}^{n} \Big(f(t, i_0) - g(t, i_0)\Big)$, we conclude that $\ell > 0$ (and, hence, $i_0 > 1$). Assume now that (4) is already proven whenever $\ell < r$ for some $r \in \mathbb{N}$. If $\ell = r$, then let $j_1 = \max L$ and $\varepsilon = \min\{f(j_1, i_0) - g(j_1, i_0), \rho\}$. Let h result from f by the diminishing transfer from (j_1, i_0) to (j_1, i_0-1) of size ε. Then $h \succ_{LO} g$ and either $\rho(h) = 0$ (if $\varepsilon = \rho(f)$) or $|L(h)| < |L(f)|$. As $\sum_{t=1}^{j} \Big(f(t, i_0) - g(t, i_0)\Big) \geqslant 0$ for all $j \in \{j_1, \ldots, n\}$ by construction, (6) is still satisfied by (1) for all $i \in \{i_0, \ldots, m\}$ so that the proof is finished by an inductive argument. **q.e.d.**

Lemma 3.4 has the following important corollary.

Corollary 3.5 *Let $n, m \in \mathbb{N}$ and $f, g \in \mathcal{F}(n, m)$ such that $f \succeq_{LO} g$. Then there exists $h \in \mathcal{F}(n, m)$ that arises from f by finitely many diminishing transfers such that $h \succeq_{LO} g$ and for all $i \in \{1, \ldots, m\}$ and all $i \in \{1, \ldots, m\}$,*

$$\begin{aligned} \widetilde{h}(j, m) - \widetilde{h}(j, i) &\geqslant \widetilde{g}(j, m) - \widetilde{g}(j, i), \\ \widetilde{h}(n, i) &= \widetilde{g}(n, i) \text{ and} \\ \widetilde{h}(j, m) &= \widetilde{g}(j, m) \end{aligned}$$

Proof: The mapping f satisfies the conditions of Lemma 3.4 for $i_0 = m$. Hence, applying the aforementioned lemma successively to $i_0 = m, \ldots, 1$ yields a mapping h that arises from f by a finite number of diminishing transfers with $h \succeq_{LO} g$ such that, by (3),

$$\widetilde{h}(j, m) - \widetilde{h}(j, i) \geqslant \widetilde{g}(j, m) - \widetilde{g}(j, i) \ \forall j \in \{j, \ldots, n-1\}, i \in \{0, \ldots, m\}.$$

Hence, applied to $i = 0$, we receive $\widetilde{h}(j, m) = \widetilde{g}(j, m)$ for all $j \in \{1, \ldots, n\}$, and, by (2), $\widetilde{h}(n, i) = \widetilde{g}(n, i)$ for all $i \in \{1, \ldots, m\}$. **q.e.d.**

So far we have shown that if $f \succeq_{LO} g$ we can obtain a function h from f by finitely many diminishing transfers that still satisfies $h \succeq_{LO} g$ and, additionally, has the same marginal distributions as g. Tchen (1980) shows that g can be obtained from an arbitrary h by correlation increasing switches if and only if g and h have coinciding marginal distributions and $h \succeq_{LO} g$. We have included a shorter proof below that proceeds by induction on m and may be illustrated as follows. We construct correlation increasing switches that successively create a new function $h \succeq_{LO} g$ such that $h(j,1) = g(j,1)$ for all $j \in \{1,\ldots,n\}$ so that we may delete the first row and apply an inductive argument. Within this row we proceed recursively by first selecting the maximal j with $f(j,1) < g(j,1)$ which we denote by j_0 and secondly selecting the minimal $j > j_0$ (denoted by j_1) such that when we subtract the aggregate weights of $\{j_0,\ldots,j\}$ of g from f this is nonnegative. We then move the difference of the aggregate weights of $\{j_0,\ldots,j_1-1\}$ via correlation increasing switches from $(j_1,1)$ to elements $(j,1)$ with $j_0 \leqslant j < j_1$ so that either j_0 or j_1 become smaller. For the precise definition of these correlation increasing switches the proof is referred to.

The following notation is useful. For $f \in \mathcal{F}$ denote

$$\mathcal{F}_f = \{g \in \mathcal{F} \mid f \succeq_{LO} g, \widetilde{g}(j,m) = \widetilde{f}(j,m), \widetilde{g}(n,i) = \widetilde{f}(n,i) \forall i \in \{1,\ldots,m\}, j \in \{1,\ldots,n\}\}.$$

This means that $\mathcal{F}_f$ contains all distributions having marginal distributions coinciding with the marginal distributions of f. Let $g \in \mathcal{F}_f$ and $f \succeq h$. Then $h \succeq g$ if and only if h arises from f by a sequence of finitely many correlation increasing switches and $g \in \mathcal{F}_h$. Moreover, denote

$$j_0(f,g) = j_0 = \max\{j \in \{1,\ldots,n\} \mid f(j,1) < g(j,1)\},$$

where $\max \emptyset = 0$ is used in this context, and

$$j_1(f,g) = j_1 = \min\left\{ j \in \{j_0+1,\ldots,n\} \,\middle|\, \sum_{t=j_0}^{j} \Big(f(t,1) - g(t,1) \Big) \geqslant 0 \right\}.$$

Note that j_1 exists because $\widetilde{g}(j_0-1,1) \geqslant \widetilde{f}(j_0-1,1)$ and $\widetilde{g}(n,1) = \widetilde{f}(n,1)$. Also note that if $j_0 = 0$, then $f(j,1) = g(j,1)$ for all $j \in \{1,\ldots,n\}$. The following technical lemma is useful.

Lemma 3.6 *Let $g \in \mathcal{F}_f$ such that $j_0(f,g) > 0$. Then there exists $f \succ h$ such that $h \succeq_{LO} g$ and $j_0(h,g) < j_0(f,g)$.*

Proof: Let $j_0 = j_0(f,g)$, $j_1 = j_1(f,g)$, and denote

$$X'(f,g) = X' = \{(j,i) \in X \mid j_0 \leqslant j < j_1, i > 1, f(j,i) > g(j,i)\}.$$

Thus $X'(f,g)$ contains all $(j,i) \in X$ in the interval between j_0 and j_1 where it is feasible to move mass into, since f has more mass here than g.

Let $\varepsilon = \sum_{t=j_0}^{j_1-1} \Big(g(t,1) - f(t,1) \Big)$. Hence, $f(j_1,1) - g(j_1,1) \geqslant \varepsilon$. As $\widetilde{f}(j,m) = \widetilde{g}(j,m)$ in particular for $j = j_0 - 1$ and $j = j_1 - 1$, the equations

$$\begin{aligned} \widetilde{f}(j_1-1,m) &= \widetilde{f}(j_0-1,m) + \textstyle\sum_{j=j_0}^{j_1-1} f(j,1) + \sum_{i=2}^{m} \sum_{j=j_0}^{j_1-1} f(j,i) \text{ and} \\ \widetilde{g}(j_1-1,m) &= \widetilde{g}(j_0-1,m) + \textstyle\sum_{j=j_0}^{j_1-1} g(j,1) + \sum_{i=2}^{m} \sum_{j=j_0}^{j_1-1} g(j,i), \end{aligned}$$

imply that $\sum_{i=2}^{m} \sum_{j=j_0}^{j_1-1} \Big(f(j,i) - g(j,i) \Big) = \varepsilon$, hence $X'(f,g) \neq \emptyset$.
By a recursive argument is suffices to construct h such that $f \succ h \succeq_{LO} g$ and either $j_0(h,g) < j_0$ or

$(j_0(h,g) = j_0$ and $j_1(h,g) < j_1)$ or $(j_0(h,g) = j_0, j_1(h,g) = j_1$, and $|X'(h,g)| < |X'|)$. For this purpose let $i' \in \{2,\dots,m\}$ be minimal such that there exists $(j,i') \in X'$. Moreover let $j' \in \{j_0,\dots,j_1-1\}$ be maximal such that $(j',i') \in X'$. Now we define $\varepsilon' = \min\{\varepsilon, f(j',i') - g(j',i')\}$ and consider $h = f_{\varepsilon'}^{(j_1,1)\leftrightarrows(j',i')}$ and verify that $h \succeq_{LO} g$. If $\varepsilon' = \varepsilon$, then we have moved ε to the element $(j',1)$ so that $\sum_{j=j_0}^{j_1-1} h(j,1) = \sum_{j=j_0}^{j_1-1} f(j,1) + \varepsilon = \sum_{j=j_0}^{j_1-1} g(j,1)$. If $j' = j_0$ and $\varepsilon = g(j_0,1) - f(j_0,1)$, we have $j_0(h,g) < j_0$. If $j' \neq j_0$ or $\varepsilon < g(j_0,1) - f(j_0,1)$, we have $j_0(h,g) = j_0$ and $j_1(h,g) < j_1$. Finally, if $\varepsilon' < \varepsilon$, then we have $j_0(h,g) = j_0, j_1(h,g) = j_1$, and $X'(h,g) = X'(f,g) \setminus \{(j',i')\}$ so that the proof is complete. **q.e.d.**

We are now able to finish the proof.

Proof of Proposition 3.3: By Corollary 3.5 we may assume that $g \in \mathcal{F}_f$. Successively applying Lemma 3.6 if necessary we may also assume that $j_0(f,g) = 0$, i.e., $f(j,1) = g(j,1)$ for all $j \in \{1,\dots,n\}$ so that the proof is finished by induction on m. **q.e.d.**

Proof of Theorem 3.1: By combining Proposition 3.2 and Proposition 3.3 the proof of our main result follows. **q.e.d.**

As mentioned earlier our constructive proof yields an algorithm. We need to follow each step in the proof, this means we have to make sets and define variables as described in the proof. There is only one place where the proof does not give us unique transfers, this is in the first step in the diminishing transfer part. Here we can decide in the algorithm in which order transfers should be made. One possibility is to choose the lexicographical order like we have done in the example below. When we follow the algorithm it is possible to track every transfer made.

4 Example

To illustrate the algorithm we give an example in the 4x4-case. Let f and g be given by Figures 4.5 and 4.6.

m				
	0.1	0.1	0.1	0.1
	0.1	0	0	0.2
	0.1	0	0.1	0.1
	0	0	0	0
				n

m				
	0.3	0.4	0.6	1
	0.2	0.2	0.3	0.6
	0.1	0.1	0.2	0.3
	0	0	0	0
				n

Figure 4.5: The probability mass function f left and $\widetilde{f}$ right.

The first part of the algorithm employs diminishing transfers to obtain equal marginal distributions. The computation of sets and values appears from the proof. An overview of the algorithm is also available in the pseudo code in Appendix A. As an example for $i_0 = 4$ we find the set K as the indices j where $\tilde{f}(j,4) - \tilde{f}(j,3) < \tilde{g}(j,4) - \tilde{g}(j,3)$. This is only true for $j = 1$, where $\tilde{f}(1,4) - \tilde{f}(1,3) = 0.3 - 0.2 = 0.1$ and $\tilde{g}(1,4) - \tilde{g}(1,3) = 0.4 - 0.2 = 0.2$. As we only have one element in K this will also be the largest element, so $j_0 = 1$. Now we know which column we need to move mass to, so now we need to check what row, $i \in \{i_0,\dots,m\}$, we can move mass from (j_0+1,i), i.e., which row $f(j_0+1,i) < g(j_0+1,i)$. Since $i_0 = 4$ this

m				
	0.2	0	0	0.2
	0	0	0.2	0
	0.1	0	0	0.1
	0.1	0	0.1	0
				n

m				
	0.4	0.4	0.7	1
	0.2	0.2	0.5	0.6
	0.2	0.2	0.3	0.4
	0.1	0.1	0.2	0.2
				n

Figure 4.6: The probability mass function g left and $\widetilde{g}$ right.

must be true for $i = 4$. Now we find $\varepsilon = \tilde{g}(1,4) - \tilde{g}(1,3) - (\tilde{f}(1,4) - \tilde{f}(1,3)) = 0.4 - 0.2 - (0.3 - 0.2) = 0.1$. This has to be divided on all elements in A. Since we only have one element in A this is quite simple. If we have more elements there will might be more ways to move the entire mass. In the example, we move as much mass as possible for the lowest indices, and stop when the entire ε is moved. This gives us a diminishing transfer in the horizontal direction. We then check whether transfers in the vertical direction is necessary, i.e., transfers that ensures that $\tilde{f}(n, i_0 - 1) = \tilde{g}(n, i_0 - 1)$ since $\tilde{f}(4,3) = 0.6 = \tilde{g}(4,3)$, no such transfers are needed. For $i_0 = 3$, $K = \emptyset$, however $\tilde{f}(4,2) = 0.3 \neq 0.4 = \tilde{g}(4,2)$. This means that $L \neq \emptyset$. We find L as the set of indices j where $\sum_{t=1}^{j} f(t, i_0) > \sum_{t=1}^{j} g(t, i_0)$ is true. $L = \{1, 2, 4\}$ since; $f(1,3) = 0.1 > 0 = g(1,3)$ and $\sum_{t=1}^{2} f(t,3) = 0.1 > 0 = \sum_{t=1}^{2} g(t,3)$ and $\sum_{t=1}^{4} f(t,3) = 0.3 > 0.2 = \sum_{t=1}^{4} g(t,3)$. We now find $\rho = \tilde{g}(4,2) - \tilde{f}(4,2) = 0.4 - 0.3 = 0.1$ and j_1 which is the largest index j where $f(j,3) > g(j,3)$ is true, $f(4,3) = 0.2 > 0 = g(4,3)$ so $j_1 = 4$. We are then able to compute $\varepsilon = \min\{f(4,3) - g(4,3), \rho\} = \min\{0.2, 0.1\} = 0.1$ and transfer this amount. Now $\rho = \rho - \varepsilon = 0.1 - 0.1 = 0$. If $\rho > 0$ after the transfer we would continue to with a new j_1 and find the new ε. In the table the steps are also shown for $i_0 = 2$ and $i_0 = 1$.

$i_0 = 4$	$K = \{1\}$, $j_0 = 1$, $A = \{4\}$, $\varepsilon = 0.1 = \varepsilon_4$	Diminishing transfer of 0.1 from $(2,4)$ to $(1,4)$
	$L = \emptyset$	
$i_0 = 3$	$K = \emptyset$	
	$L = \{1,2,4\}$, $\rho = 0.1$, $j_1 = 4$, $\varepsilon = 0.1$	Diminishing transfer of 0.1 from $(4,3)$ to $(4,2)$
$i_0 = 2$	$K = \emptyset$	
	$L = \{3,4\}$, $\rho = 0.2$, $j_1 = 4$, $\varepsilon = 0.1$	Diminishing transfer of 0.1 from $(4,2)$ to $(4,1)$
	$L = \{3\}$, $\rho = 0.1$, $j_1 = 3$, $\varepsilon = 0.1$	Diminishing transfer of 0.1 from $(3,2)$ to $(3,1)$
$i_0 = 1$	$K = \{3\}$, $j_0 = 3$, $A = \{1,3\}$, $\varepsilon = 0.1 = \varepsilon_1$	Diminishing transfer of 0.1 from $(4,1)$ to $(3,1)$

After diminishing transfers f looks like Figure 4.7. When we have secured that the marginal distributions are equal we will continue with the second part where we employ correlation increasing switches in order to obtain identical probability mass functions. Here $i_0 = i$ means that the row i from the full probability mass functions is the bottom row in the present step. This means that in our 4x4-example having $i_0 = 4$ indeed means that we only consider the uppermost row from the full probability mass functions, because the rest are already identical. We start with $i_0 = 1$. Here we find $j_0 = \max\{j \in \{1, \dots, n\} | f(j,1) < g(j,1)\}$. We have that $f(1,1) = 0 < 0.1 = g(1,1)$, $f(2,1) = 0 = g(2,1)$, $f(3,1) = 0.2 > 0.1 = g(3,1)$ and $f(4,1) = 0 = g(4,1)$, which means that $j_0 = 1$. Then we find

m	0.2	0	0.1	0.1
	0.1	0	0	0.1
	0.1	0	0	0.1
	0	0	0.2	0
			n	

m	0.4	0.4	0.7	1
	0.2	0.2	0.4	0.6
	0.1	0.1	0.3	0.4
	0	0	0.2	0.2
				n

Figure 4.7: The new probability mass function f left and $\widetilde{f}$ right.

$j_1 = \min\{j \in \{j_0+1, \dots, n\} | \sum_{t=j_0}^{j} \Big(f(t,1) - g(t,1)\Big) \geq 0\}$, since $\sum_{t=1}^{2} \Big(f(t,1) - g(t,1)\Big) = -0.1$ and $\sum_{t=1}^{3} \Big(f(t,1) - g(t,1)\Big) = 0$ we have that $j_1 = 3$. We have to compute the set $X' = \{(j,i) \in X | j_0 \leq j < j_1, i > i_0, f(j,i) > g(j,i)\}$ this gives us $X' = \{(1,3)\}$. This means that when we have to chose i' minimal such that $(j,i') \in X'$ and j' maximal such that $(j',i') \in X'$ we will have $(j',i') = (3,1)$. We can compute $\varepsilon = \sum_{t=j_0}^{j_1-1} \Big(g(t,i_0) - f(t,i_0)\Big) = (g(1,1) - f(1,1)) + (g(2,1) - f(2,1)) = (0.1-0) + (0-0) = 0.1$ and $\varepsilon' = \min\{\varepsilon, f(j',i') - g(j',i')\} = \min\{\varepsilon, f(1,3) - g(1,3)\} = \min\{0.1, 0.1-0\} = 0.1$. So now we have identified a correlation increasing switch. Since the row $i = 1$ is identical in f and g we are done with this row. So we can discard it from the algorithm and look on the remaining rows. In the table the rows $i_0 = 2$ and $i_0 = 4$ are blank, this is because these rows are already similar, which means it is not necessary to employ any correlation increasing switches.

$i_0 = 1$	$j_0 = 1$, $j_1 = 3$, $i' = 3$, $j' = 1$, $\varepsilon = 0.1$, $\varepsilon' = 0.1$ Correlation increasing switch of 0.1 from $(3,1)$ to $(1,1)$ and from $(1,3)$ to $(3,3)$
$i_0 = 2$	
$i_0 = 3$	$j_0 = 3$, $j_1 = 4$, $i' = 4$, $j' = 3$, $\varepsilon = 0.1$, $\varepsilon' = 0.1$ Correlation increasing switch of 0.1 from $(4,3)$ to $(3,3)$ and from $(3,4)$ to $(4,4)$
$i_0 = 4$	

5 Discussion

As mentioned earlier, Theorem 3.1 was recently proven by Meyer and Strulovici (2015) and Müller (2013). However, our approach is different. It particular, our proof is constructive and yields an algorithm. This algorithm is useful in several ways. The most important feature is that it allows us to decompose domination into diminishing transfers and correlation increasing switches. In order words, we can disentangle domination into welfare deteriorations and inequality increases. We can for example measure how much mass is moved, e.g., the amount of mass moved times the distance it is moved, using ℓ_1 norm as distance. Then the algorithm tells us how much the distributions differ, and, in particular, we can decompose the domination, to see how much mass is moved by diminishing transfers and how much mass is moved by correlation increasing switches. Although the system of diminishing transfers and correlation increasing switches leading from one distribution to another is generally not uniquely determined, it is important to notice that the mass we have to move with diminishing transfers to obtain the same marginal distributions, is always the same.[7] Thus, in terms of the measure of distance of mass displacement through diminishing transfers, the welfare differences are uniquely determined and hence meaningfully measured for comparison purposes

References

Aaberge, R., and A. Brandolini (2014): "Multidimensional poverty and inequality", *Bank of Italy Temi di Discussione (Working Paper)*, 976.

Arndt, C., R. Distante, M. A. Hussain, L. P. Østerdal, P. L. Huong, and M. Ibraimo (2012): "Ordinal Welfare Comparisons with Multiple Discrete Indicators: A First Order Dominance Approach and Application to Child Poverty", *World Development*, 40, 2290 – 2301.

Atkinson, A., and F. Bourguignon (1982): "The comparison of multi-dimensioned distributions of economic status", *Review of Economics Studies*, 49, 183–201.

Bourguignon, F., and S. R. Chakravarty (2003): "The measurement of multidimensional poverty", *The Journal of Economic Inequality*, 1, 25–49.

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

Duclos, J.-Y., D. E. Sahn, and S. D. Younger (2006): "Robust multidimensional poverty comparisons", *Economic Journal*, 116, 943–968.

——— (2007): "Robust multidimensional poverty comparisons with discrete indicators of well", in *Inequality and poverty re-examined*, ed. by S. P. Jenkins, and J. Micklewright. Oxford University Press.

Gravel, N., P. Moyes, and B. Tarroux (2009): "Robust international comparisons of distributions of disposable income and regional public goods", *Economica*, 76(303), 432–461.

Gravel, N., and A. Mukhopadhyay (2010): "Is India better off today than 15 years ago? A robust multidimensional answer", *The Journal of Economic Inequality*, 8, 173–195.

Kamae, T., U. Krengel, and G. O'Brien (1977): "Stochastic inequalities on partially ordered spaces", *The Annals of Probability*, 5, 899–912.

Levy, H. (1992): "Stochastic dominance and expected utility: survey and analysis", *Management Science*, 38, 555–593.

Meyer, M., and B. Strulovici (2015): "Beyond Correlation: Measuring Interdependence Through Complementarities", *Unpublished*.

Moyes, P. (2012): "Comparisons of heterogeneous distributions and dominance criteria", *Journal of Economic Theory*, 147(4), 1351–1383.

[7] Consider distributions f, g and h. Suppose that g can be obtained from f by a sequence of diminishing transfers, and similarly, h can be obtained from f by (another) sequence of diminishing transfers. Moreover, assume that g and h have identical partial marginal distributions. We have to prove that the transfer system leading from g to f "moves the same amount of mass" as the system obtaining h from f. First realize that the result is true for the case that $m = 1$, i.e., the one-dimensional case, because the partial marginal distributions unique defined the distribution. If $m > 1$, then we know exactly how much mass has to be moved from right to left. Similarly, by the one-dimensional case, we also know exactly how much mass has to be moved down. Therefore we know exactly how much mass to move.

Müller, A. (2013): "Duality theory and transfers for stochastic order relations", in *Stochastic Orders in Reliability and Risk in Honor of Professor Moshe Shaked*, ed. by H. Li, and X. Li, Lecture Notes in Statistics 208, pp. 41 – 57, New York. Springer.

Müller, A., and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*. John Wiley and Sons.

Muller, C., and A. Trannoy (2011): "A dominance approach to the appraisal of the distribution of well-being across countries", *Journal of Public Economics*, 95(3), 239–246.

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

Range, T. M., and L. P. Østerdal (2013): "Checking bivariate first order dominance", Discussion paper, Department of Business and Economics, University of Southern Denmark.

Shaked, M., and J. Shanthikumar (2007): *Stochastic Orders*. Springer.

Silvapulle, M. J., and P. K. Sen (2011): *Constrained statistical inference: Order, inequality, and shape constraints.*, vol. 912. Wiley-Interscience.

Sonne-Schmidt, C., F. Tarp, and L. P. Østerdal (2015): "Ordinal bivariate inequality: concepts and application to child deprivation in Mozambique", *Review of Income and Wealth*, in press. DOI: 10.1111/roiw.12183.

Sriboonchita, S., S. Dhompongsa, W. K. Wong, and H. T. Nguyen (2009): *Stochastic dominance and applications to finance, risk and economics*. Chapman & Hall/CRC.

Strassen, V. (1965): "The existence of probability measures with given marginals", *The Annals of Mathematical Statistics*, 36, 423–439.

Tchen, A. H. (1980): "Inequalities for distributions with given marginals", *The Annals of Probability*, pp. 814–827.

Tsui, K.-y. (1999): "Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation", *Social Choice and Welfare*, 16, 145–157.

A Pseudo code

Data: Two probability mass functions f and g

Result: If there is dominance a series of diminishing transfers and correlation increasing switches is returned

```
Diminishing bilateral transfers:
for i_0 = m to 1 do
    Step 1:
    K = {{j \in {1,...,n}} | \tilde{f}(j,m) - \tilde{f}(j,i_0 - 1) < \tilde{g}(j,m) - \tilde{g}(j,i_0 - 1)}
    j_0=max K;
    if j_0 = n then
        Print "There is no dominance";
        exit;
    end
    while K \neq \emptyset do
        A = {i \in {i_0,...,m} | f(j_0 + 1, i) > g(j_0 + 1, i)};
        \varepsilon = \tilde{g}(j_0,m) - \tilde{g}(j_0,i_0 - 1) - (\tilde{f}(j_0,m) - \tilde{f}(j_0,i_0 - 1));
        for i = 1 to m do
            \varepsilon_i = min{f(j_0 + 1, i) - g(j_0 + 1, i), \varepsilon - \sum_{t=i_0}^{m} \varepsilon_i};
        end
        foreach i \in A do
            Transfer \varepsilon_i from (j_0 + 1, i) to (j_0, i);
        end
        Remove j_0 from K;
    end
    Step 2:
    \rho = \tilde{g}(n, i_0 - 1) - \tilde{h}(n, i_0 - 1);
    L = {j \in {1,...,n} | \sum_{t=1}^{j} f(t,i_0) > \sum_{t=1}^{j} g(t,i_0)};
    while L \neq \emptyset do
        j_1 = max{j \in L | f(j,i_0) > g(j,i_0)};
        \varepsilon = min{f(j_1,i_0) - g(j_1,i_0), \rho};
        Transfer \varepsilon from (j_1,i_0) to (j_1,i_0 - 1);
        \rho = \rho - \varepsilon;
    end
end
Correlation increasing switches:
for i_0 = 1 to m - 1 do
    while f(j,i_o) \neq g(j,i_o) \forall j \in {1,...,n} do
        j_0 = max{j \in {1,...,n} | f(j,i_0) < g(j,i_0)};
        j_1 = min{ j \in {j_0 + 1,...,n} | \sum_{t=j_0}^{j} ( f(t,i_0) - g(t,i_0) ) \geq 0 };
        if j_1 does not exist then
            Print "There is no dominance";
            exit;
        end
        \varepsilon = \sum_{t=j_0}^{j_1 - 1} ( g(t,i_0) - f(t,i_0) );
        while \varepsilon > 0 do
            X' = {j,i} \in X | j_0 \leq j < j_1, i > i_0, f(j,i) > g(j,i)};
            i' = min{i \in {i_0 + 1,...,m} | \exists (j,i') \in X'};
            j' = max{j \in {j_0,...,j_1 - 1} | \exists (j',i') \in X'};
            \varepsilon' = min{\varepsilon, f(j',i') - g(j',i')};
            Transfer \varepsilon' from (j_1,i_0) to (j',i_0) and \varepsilon' from (j',i') to (j_1,i');
            \varepsilon = \varepsilon - \varepsilon';
        end
    end
end
```

The algorithm described in the pseudo code is slightly different from the algorithm that can be derived from the proof. The main difference is that the algorithm, which the proof yields, only takes two functions where dominance is present as input. The algorithm in the pseudo code will take any two functions and return a series of diminishing transfers and correlation increasing switches if lower orthant dominance occur. If there is no dominance the algorithm will terminate with the message of no dominance. The first place, where we insert this stop-message is in line 6-9 in the pseudo code. Here we check whether $j_0 = n$ if this is the case there will be no dominance, because then the conditions for lower orthant dominance is not fulfilled. The second place where the stop-message is needed is in line 36-38, we check whether j_1 exists, if this is not the case the conditions for lower orthant dominance will not be fulfilled.

It is simple to introduce variables that accumulate mass moved by diminishing mass transfers and correlation increasing switches respectively. Furthermore it is done in constant time, thus it will not affect the complexity. The complexity of the diminishing transfer part of the algorithm is $O(n \log n \; m)$ and the complexity of the correlation increasing switch part is $O(n^2m^2)$. The complexity of the entire algorithm will thus be $O(n^2m^2)$. This may be improved by effective implementation.