博士論文　2019年度

Social Choice and Measurement

慶應義塾大学 大学院 経済学研究科

河田　陽向

Abstract

This thesis consists of four essays on social choice and measurement. I study various "aggregation rules" such as voting rules, electoral systems, aggregation formulas of Human Development Index, and measures of population ageing. These are functions that aggregate "many" to "one". For example, voting rules aggregate different individual preferences and choose one alternative from others. Similarly, a measure of population ageing is a function that maps individual ages to one real number that indicates the level of population ageing in a society.

In Chapter 1, I propose a voting rule based on *cosine similarity*. Cosine similarity is a commonly used similarity measure in computer science. I apply this similarity measure to define a voting rule, namely, the *cosine similarity rule*. This rule selects a social ranking that maximizes cosine similarity between the social ranking and a given preference profile. Our main finding is that the cosine similarity rule in fact coincides with the Borda rule.

In Chapter 2, I study electoral systems in representative democracies. I define an electoral system as a function that maps each preference profile to the distribution of seats among political parties in the congress or legislative chambers. My purpose is to search for electoral systems in which the distribution of seats in the congress can appropriately reflect preferences of the people in the nation. I introduce two consistency conditions for electoral systems, *Condorcet consistency* and *Borda consistency*. I first present a paradox of single-member district systems, namely, the *Loser Dominance Paradox*, which exhibits difficulty in the consistent aggregation of preferences of the people. Next, I show that single-member district systems and simple proportional representation systems violate both consistency conditions. Finally, I propose a new electoral system, namely, the *Borda proportional representation system* and show that it satisfies both consistency conditions.

In Chapter 3, I propose a new approach for multidimensional evaluation when achievements in different dimensions are not easily comparable. Our approach can be applied to measurements of well-being based on capability approach such as human development or multidimensional poverty. In measurements of such things, we should respect (i) monotonicity to each achievement and (ii) incomparability across different dimensions. However, any method currently in use does not respect (i) or (ii). I introduce a new axiom *dimensional*

independence that captures incomparability across different dimensions. Then, I propose a new methods for multidimensional evaluation and show that our methods satisfy both of *monotonicity* and *dimensional independence*. Moreover, in a certain class of methods, I find a unique method that satisfies monotonicity, dimensional independence, and *minimal lower boundedness*. I apply this method for measurement of human development and compute a new *human development indices* of 188 countries.

In Chapter 4 (co-authored with Yuta nakamura and Noriaki Okamoto), we study the measurement of population ageing. Population ageing is one of the most serious problems in many developed countries. The level of population ageing is often measured by "usual" measures such as the share of the older population, mean age, median age, and the dependency ratio. However, these measures violate elementary properties for measuring population ageing. We propose a new measure of population ageing that overcomes drawbacks of the measures currently in use. We introduce a new condition called the *working age principle*, which is a sensitivity condition to thickness of the working age population. Our measure is the only measure that satisfies *monotonicity*, *continuity*, *separability*, *normalization*, and the *working age principle*.

Acknowledgments

I would like to express my sincerest gratitude to my advisor, Professor Toyotaka Sakai, for his continued encouragement and patience. He taught me how to choose research themes that I should study and how to think about it in the very early stage of my career. Regular meetings with him have been a good guide for me to point the right direction to go on.

I am also deeply indebted to Professors Takako Fujiwara-Greve, Toru Hokari, Shinsuke Nakamura for their invaluable advice. They gave me a lot of helpful comments and insightful questions to help me improve the quality of this thesis.

I have learned many things from discussions with Noriaki Okamoto and Yuta Nakamura. They are also co-authors of Chapter 4. I am very honored to be a co-author and friend with them.

I am also grateful for comments and discussions from other faculties of Keio University and participants at seminars or conferences at Keio University, the University of Tokyo, Hitotsubashi University, Nagoya University, Niigata University, Waseda University, Fukuoka University, Sophia University, Aoyama-Gakuin University, Seoul National University, and the University of York.

Contents

Chapter 1

Cosine Similarity and the Borda Rule

1.1 Introduction

Cosine similarity is a commonly used similarity measure in computer science. It has a variety of applications such as document clustering, search engines, and face verification.[*1] We apply this similarity measure to define a voting rule, namely, the *cosine similarity rule*. The rule selects a social ranking that is closest to a given preference profile when measured by cosine similarity. Our main finding is that the cosine similarity rule in fact coincides with the Borda rule.

The Borda rule is one of the most important voting rules in social choice theory, which is introduced by Jean-Charles de Borda (1784). It is known that Borda's choice rule has many desirable properties such as maximization of the average share of votes in pairwise comparison (Black 1976, Coughlin 1979), closest proximity to unanimous agreement (Sen 1977; Farkas and Nitzan 1979), avoidance of many paradoxes observed in positional rules (Saari 1989), and avoidance of the Condorcet loser (Fishburn and Gehrlein 1976, Okamoto and Sakai 2013). In particular, Young (1974) characterizes Borda's choice rule by a set of desirable properties: neutrality, consistency, faithfulness, and cancellation. On the other hand, there are relatively a few studies on Borda's *ranking* rule. In this chapter, we focus on the Borda rule as a ranking rule. Our result provides a rationale for the use of the Borda rule based on cosine similarity. The Borda rule selects the ranking closest to a given preference profile when measured by cosine similarity.

This result is parallel to the better-known characterization of the *Condorcet rule*. In his seminal work, Kemeny (1959) searches for desirable ranking rules based on distance from a given preference profile. He defines a metric that measures the distance between

[*1]Singhal (2001) briefly explains how to apply cosine similarity to measure the similarity between two text documents. Cosine similarity is also used as a basis of search engines (see, e.g., Bayardo, Ma, and Srikant 2007). An application for face verification system is a recent interesting example of applications of cosine similarity (Nguyen and Bai 2010).

two rankings, so called *Kemeny distance*. Then he proposes the ranking rule that selects a ranking minimizing the sum of Kemeny distances between the social ranking and each voter's preference. Surprisingly, it is the unique ranking rule that satisfies neutrality, consistency, and the Condorcet criterion (Young and Levenglick 1978). Moreover, Young (1988) argues that what Condorcet (1785) had in mind was in fact the maximum likelihood method, and he shows that Kemeny's rule coincides with the maximum likelihood method. Summarizing these results, Condorcet's rule selects a ranking that minimizes the sum of Kemeny distances from voters' preferences. This characterization of the Condorcet rule is parallel to our result that Borda's rule selects a ranking that maximizes the sum of cosine similarities from voter's preferences.[*2] As Saari (2006) notes, "Condorcet or Borda, which is better?" is the two-century old question in social choice theory. Now the choice between these two rules can be attributed to the choice between the Kemeny minimization and the cosine similarity maximization. If one chooses the former, he is recommended to use Condorcet's rule, and if one chooses the latter, he is recommended to use Borda's rule, as suggested by our main result.

This chapter is organized as follows. In Section 1.2 we introduce definitions. In Section 1.3 we show the equivalence between the Borda rule and the cosine similarity rule. In Section 1.4 we discuss two topics: an implication of restricting the range to the set of linear orderings and a relationship between the cosine similarity rule and scoring rules. Section 1.5 concludes this chapter.

1.2 Definitions

Let $I = \{1, 2, \ldots, n\}$ be a finite set of voters and $A = \{a_1, a_2, \ldots, a_m\}$ a finite set of alternatives. Let $\mathcal{R} \subset A \times A$ be the set of *complete* and *transitive* binary relations on A and $\mathcal{P} \subset A \times A$ the set of *complete*, *transitive* and *anti-symmetric* binary relations on A.[*3] Each voter $i \in I$ has a preference relation $\succsim_i \in \mathcal{P}$ on A. A *preference profile* is a list of preference relations

$$\succsim \equiv (\succsim_i)_{i \in I} \in \mathcal{P}^n.$$

[*2] Saari and Merlin (2000) conduct a fairly complete analysis of the Kemeny ranking rule. Particularly, they reveal a deep relationship between the Borda ranking and the Kemeny ranking.

[*3] A binary relation $\succsim$ is *complete* if for any $a, b \in A$, $a \succsim b$ or $b \succsim a$. It is *transitive* if for any $a, b, c \in A$, $[a \succsim b$ and $b \succsim c]$ implies $a \succsim c$. It is *anti-symmetric* if for any $a, b \in A$, $[a \succsim b$ and $b \succsim a]$ implies $a = b$.

A *ranking rule* is a function F that maps each preference profile $\succsim \in \mathcal{P}^n$ to a social ranking $F(\succsim) \in \mathcal{R}$.

For each alternative $a \in A$ and each preference $\succsim_i \in \mathcal{P}$, let

$$r_a(\succsim_i) \equiv |\{a' \in A : a \succsim_i a'\}|$$

be the *inverse ranking* of $a \in A$ in $\succsim_i$. For example, $r_a(\succsim_i) = 1$ means that a is the worst alternative for i, and $r_a(\succsim_i) = m$ means that a is most preferred by i. For each $\succsim_i \in \mathcal{P}$, define

$$r(\succsim_i) \equiv (r_{a_1}(\succsim_i), r_{a_2}(\succsim_i), \ldots, r_{a_m}(\succsim_i)) \in \mathbb{N}^m.$$

We call $r(\succsim_i)$ the *rank expression* of $\succsim_i$. A preference $\succsim_i$ and its rank expression $r(\succsim_i)$ have the same information. The *Borda score* of $a \in A$ in $\succsim$ is given by

$$S(a, \succsim) \equiv \sum_{i \in I} r_a(\succsim_i).$$

The following ranking rule is proposed by Borda (1784).

Definition 1 (Borda rule)**.** The *Borda rule* is the ranking rule F^B such that for each $\succsim \in \mathcal{P}^n$ and each $a, b \in A$,

$$a \; F^B(\succsim) \; b \text{ if and only if } S(a, \succsim) \geq S(b, \succsim).$$

Next, we define a similarity measure that plays a key role in our analysis.

Definition 2 (Cosine similarity)**.** For each vector $x, y \in \mathbb{R}^m_{++}$, *cosine similarity* between x and y is

$$C(x, y) \equiv \frac{x \cdot y}{\|x\| \|y\|},$$

where $\|x\|$ is the Euclidean norm of x, and $x \cdot y$ denotes the inner product between x and y.

Cosine similarity is a commonly used similarity measure between two vectors. By definition, $C(x, y) = \cos \theta_{x,y}$, where $\theta_{x,y}$ is the angle between x and y. Therefore, $C(x, x) = \cos 0 = 1$ and $C(x, y) \leq 1$ for all $x, y \in \mathbb{R}^m_{++}$. If the value of $C(x, y)$ is close to 1, we say that the vectors x and y are similar.

For example, consider the following four vectors x, y, z and w.

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \qquad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 2 \end{pmatrix}, \qquad z = \begin{pmatrix} 10 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad w = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}.$$

We have

$$\begin{aligned} C(x, y) &= \frac{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 2}{\sqrt{1^2 + 2^2 + 3^2 + 4^2}\sqrt{1^2 + 2^2 + 3^2 + 2^2}} \fallingdotseq .9467, \\ C(x, z) &= \frac{1 \cdot 10 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 1}{\sqrt{1^2 + 2^2 + 3^2 + 4^2}\sqrt{10^2 + 1^2 + 1^2 + 1^2}} \fallingdotseq .3418, \\ C(x, w) &= \frac{1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6 + 4 \cdot 8}{\sqrt{1^2 + 2^2 + 3^2 + 4^2}\sqrt{2^2 + 4^2 + 6^2 + 8^2}} = 1. \end{aligned}$$

Note that cosine similarity measures the similarity of orientations of two vectors and is independent from their lengths. Using cosine similarity and rank expressions of preferences, we can calculate the similarity between two preferences.

Definition 3 (Cosine similarity between two preferences)**.** For each $\succsim_i$ and $\succsim_j \in \mathcal{P}$, the *cosine similarity* between $\succsim_i$ and $\succsim_j$ is $C(r(\succsim_i), r(\succsim_j))$.

A social ranking may be a weak ordering that cannot be associated with any rank expression r, so we introduce the following notation. For each vector $x \in \mathbb{R}^m_{++}$, we call $R(x) \in \mathcal{R}$ the *ranking expressed by* x if for each $k, \ell \in \{1, 2, \ldots, m\}$,

$$a_k \; R(x) \; a_\ell \Longleftrightarrow x_k \geq x_\ell.$$

We will define a new voting rule that selects the social ranking closest to a given preference profile when measured by this similarity measure.

Definition 4 (Cosine similarity rule)**.** The *cosine similarity rule* is the ranking rule F^C such that for each $\succsim \in \mathcal{P}^n$,

$$F^C(\succsim) = R(x), \text{ where } x \text{ maximizes } \sum_{i \in I} C(\,\cdot\,, r(\succsim_i)).^{*4}$$

This rule selects a social ranking that maximizes the sum of cosine similarities between the social ranking and each voter's ranking[*5].

1.3 Equivalence theorem

We are now in a position to state our main result. We show that the cosine similarity rule coincides with the Borda rule.

Theorem 1. *For each $\succsim \in \mathcal{P}^n$,*

$$F^C(\succsim) = F^B(\succsim).$$

In other words, for each preference profile, the Borda rule selects the social ranking closest to the preference profile when measured by cosine similarity. As already mentioned, the Kemeny rule, which selects the social ranking closest to a given preference profile when measured by Kemeny distance, coincides with the Condorcet rule. Figure 1 illustrates relations between the Kemeny rule, the Condorcet rule, the cosine similarity rule, and the Borda rule. While Condorcet's rule minimizes Kemeny distance, Borda's rule maximizes cosine similarity.

Proof. Take any $\succsim \in \mathcal{P}^n$. Assume that $x \in \mathbb{N}^m$ maximizes

$$\sum_{i \in I} C(\,\cdot\,, r(\succsim_i)).$$

Step 1 (Existence of α). We show that there exists $\alpha \in \mathbb{R}_{++}$ such that $x_k = \alpha \cdot S(a_k, \succsim)$ for

[*4]The cosine similarity rule is a well-defined (single-valued) function. Indeed, a vector x is not uniquely determined by $\succsim$ but we will show that $R(x)$ is uniquely determined by $\succsim$ in the proof of Theorem 1.

[*5]In the definition of the cosine similarity rule, we use vector expression x. This point is different from that of Young and Levenglick (1978), but it is not essential for our results. In Section 1.4.1, we define a cosine similarity rule without using vector expression and consider only linear orderings, as Young and Levenglick (1978) do.

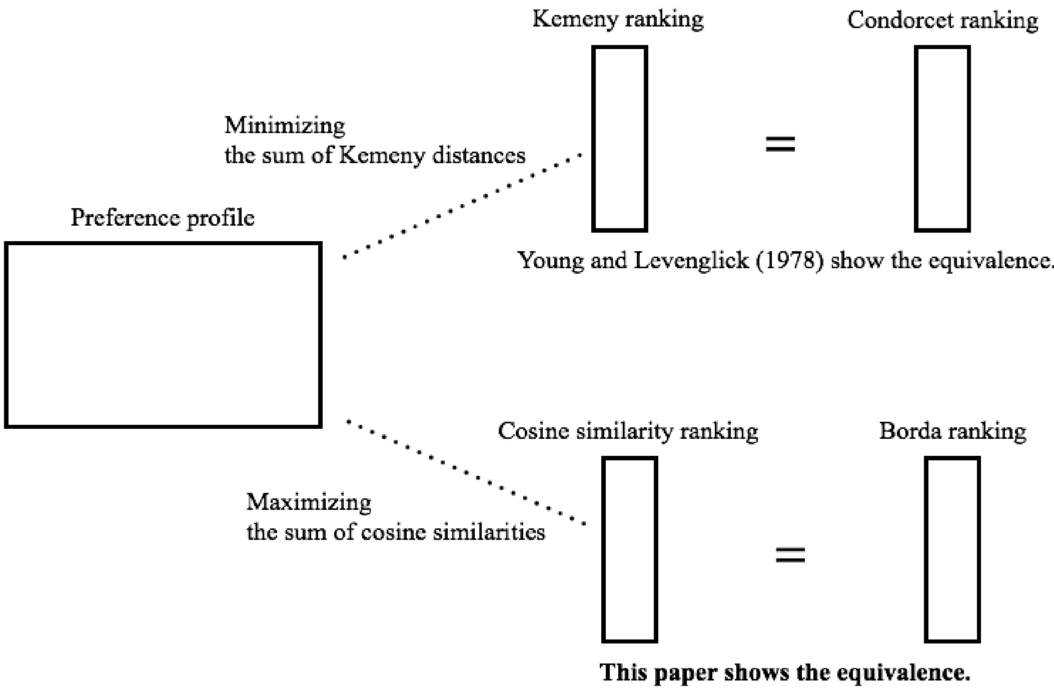

Figure 1: Borda's rule and Condorcet's rule

all $k = 1, 2, \ldots, m$. By the definition of cosine similarity,

$$\begin{aligned}\sum_{i \in I} C(x, r(\succsim_i)) &= \sum_{i \in I} \frac{x \cdot r(\succsim_i)}{\|x\| \|r(\succsim_i)\|} \\ &= \sum_{i \in I} \frac{\sum_{j=1}^{m} x_j r_{a_j}(\succsim_i)}{\|x\| \|r(\succsim_i)\|}.\end{aligned}$$

Let $\|r\| \equiv \|r(\succsim_i)\| = \sqrt{1^2 + 2^2 + \cdots + m^2}$ for all $i \in I$. By the commutative property of addition,

$$\begin{aligned}\sum_{i \in I} \frac{\sum_{j=1}^{m} x_j r_{a_j}(\succsim_i)}{\|x\| \|r(\succsim_i)\|} &= \frac{1}{\|r\|} \sum_{i \in I} \frac{\sum_{j=1}^{m} x_j r_{a_j}(\succsim_i)}{\|x\|} \\ &= \frac{1}{\|r\|} \sum_{j=1}^{m} \frac{x_j \sum_{i \in I} r_{a_j}(\succsim_i)}{\|x\|}.\end{aligned}$$

Since $\sum_{i\in I}^m r_{a_j}(\succsim_i) = S(a_j, \succsim)$,

$$\frac{1}{\|r\|}\sum_{j=1}^m \frac{x_j \sum_{i\in I} r_{a_j}(\succsim_i)}{\|x\|} = \frac{1}{\|r\|}\sum_{j=1}^m \frac{x_j S(a_j,\succsim)}{\|x\|}$$
$$= \frac{\|(S(a_j,\succsim))_{j=1}^m\|}{\|r\|}\frac{\sum_{j=1}^m x_j S(a_j,\succsim)}{\|x\|\|(S(a_j,\succsim))_{j=1}^m\|}$$
$$= \frac{\|(S(a_j,\succsim))_{j=1}^m\|}{\|r\|}\frac{x\cdot (S(a_j,\succsim))_{j=1}^m}{\|x\|\|(S(a_j,\succsim))_{j=1}^m\|},$$

where $(S(a_j,\succsim))_{j=1}^m$ is the vector $(S(a_1,\succsim),\ldots,S(a_m,\succsim),)\in\mathbb{N}^m$. By the definition of cosine similarity,

$$\frac{x\cdot (S(a_j,\succsim))_{j=1}^m}{\|x\|\|(S(a_j,\succsim))_{j=1}^m\|} = C(x, S(a_j,\succsim))_{j=1}^m) \leq 1.$$

Therefore,

$$\sum_{i\in I} C(x, r(\succsim_i)) \leq \frac{\|(S(a_j,\succsim))_{j=1}^m\|}{\|r\|}. \tag{1.1}$$

Since $C((S(a_j,\succsim))_{j=1}^m, (S(a_j,\succsim))_{j=1}^m) = 1$, by *scale invariance property* of cosine similarity[*6], there exists some positive real number α such that

$$\sum_{i\in I} C(x, r(\succsim_i)) = \frac{\|(S(a_j,\succsim))_{j=1}^m\|}{\|r\|} \Longleftrightarrow x = \alpha\cdot (S(a_j,\succsim))_{j=1}^m. \tag{1.2}$$

Step 2 (Completing the proof). Let us complete our proof. By (1.2), since $\alpha > 0$,

$$x_k \geq x_j \Longleftrightarrow S(a_k,\succsim) \geq S(a_j,\succsim) \qquad \forall a_k, a_j \in A.$$

By the definition of $R(x)$,

$$x_k \geq x_j \Longleftrightarrow a_k\ R(x)\ a_j \qquad \forall a_k, a_j \in A.$$

[*6]Cosine similarity is scale invariant, that is, for each $x, y \in \mathbb{R}^m_{++}$ and for each $\alpha > 0$, $C(\alpha\cdot x, y) = C(x, y)$.

Therefore,

$$S(a_k, \succsim) \geq S(a_j, \succsim) \Longleftrightarrow a_k \ R(x) \ a_j \qquad \forall a_k, a_j \in A. \tag{1.3}$$

This implies that the ranking $R(x) = F^C(\succsim)$ is uniquely determined by $\succsim$. Moreover, by (1.3) and the definition of the Borda ranking, $F^B(\succsim) = R(x)$. Therefore, $F^C(\succsim) = F^B(\succsim)$. □

1.4 Discussion

1.4.1 Linear orderings and another equivalence

Since there are cases in which two alternatives obtain the same Borda scores, it is natural to assume that the range of ranking rules is the set of weak orderings. Indeed, in the previous section, we establish the equivalence between the weak Borda rankings and the weak cosine similarity rankings. However, Young and Levenglick (1978) show the equivalence between the Kemeny rankings and the Condorcet rankings on the restricted range of linear orderings. By considering only linear orderings, we can relate Young and Levenglick's (1978) analysis with our analysis. In this section, we show the equivalence between the linear Borda rule and the linear cosine similarity rule.

Definition 5 (Linear Borda rule)**.** The *linear Borda rule* is a correspondence $F^{LB} : \mathcal{P}^n \twoheadrightarrow \mathcal{P}$ such that for each $\succsim \in \mathcal{P}^n$,

$$F^{LB}(\succsim) = \{P \in \mathcal{P} : a_j \ P \ a_k \Longrightarrow S(a_j, \succsim) \geq S(a_k, \succsim) \ \forall a_j, a_k \in A\}. \tag{1.4}$$

For example, suppose $S(a, \succsim) = 11$, $S(b, \succsim) = 14$, $S(c, \succsim) = 11$, and $S(d, \succsim) = 4$. Then the set of linear Borda rankings is

$$F^{LB}(\succsim) = \{P, P'\},$$

where

$$b \ P \ c \ P \ a \ P \ d \quad \text{and} \quad b \ P' \ a \ P' \ c \ P' \ d.$$

Definition 6 (Linear cosine similarity rule)**.** The *linear cosine similarity rule* is a correspondence $F^{LC} : \mathcal{P}^n \twoheadrightarrow \mathcal{P}$ such that for each $\succsim \in \mathcal{P}^n$,

$$F^{LC}(\succsim) = \{P \in \mathcal{P} : P \text{ maximizes } \sum_{i \in I} C(r(P), r(\succsim_i))\}.$$

The linear cosine similarity rule can be defined without using vector expressions, so its definition is simpler than that of the weak ordering cosine similarity rule.

Theorem 2. *For each* $\succsim \in \mathcal{P}^n$,

$$F^{LB}(\succsim) = F^{LC}(\succsim).$$

Before giving the proof, we offer a lemma. The next lemma is known as the rearrangement inequality. Since it is nontrivial, we write its proof in Appendix 1.[*7]

Lemma 1. *For any* $x, y \in \mathbb{R}^m$ *with* $x_1 \leq x_2 \leq \cdots \leq x_m$*,* $y_1 \leq y_2 \leq \cdots \leq y_m$ *and any permutation* σ *on* $\{1, 2, \ldots, m\}$*,*

$$\sum_{k=1}^{m} x_k y_k \geq \sum_{k=1}^{m} x_k y_{\sigma(k)}.$$

Proof. See Appendix 1. □

Proof of Theorem 2. Consider any $\succsim \in \mathcal{P}^n$. Let $s \in \mathbb{N}^m$ be such that $s_j = S(a_j, \succsim)$ for each $a_j \in A$. First, we show that $F^{LC}(\succsim) \subset F^{LB}(\succsim)$. Take any $P \in F^{LC}(\succsim)$. By an argument similar to Step 1 in the proof of Theorem 1,

$$\begin{aligned}\sum_{i \in I} C(r(P), r(\succsim_i)) &= \frac{\|s\|}{\|r\|} \frac{r(P) \cdot s}{\|r(P)\| \|s\|} \\ &= \frac{r_1(P)s_1 + r_2(P)s_2 + \cdots + r_m(P)s_m}{\|r\| \|r(P)\|}.\end{aligned}$$

Since P is a linear order, $\|r(P)\| = \sqrt{\frac{1}{6}m(m+1)(2m+1)}$, which is independent of the choice of P. Therefore,

$$P \text{ maximizes } \sum_{i \in I} C(r(\cdot), r(\succsim_i)) \iff P \text{ maximizes } r_1(\cdot)s_1 + r_2(\cdot)s_2 + \cdots + r_m(\cdot)s_m. \quad (1.5)$$

[*7] Another proof can be seen in Hardy, Littlewood, and Pólya (1952).

Since $P \in F^{LC}(\succsim)$, P maximizes $\sum_{i \in I} C(r(\cdot), r(\succsim_i))$, and by (1.5), P is a maximizer of $r_1(\cdot)s_1 + r_2(\cdot)s_2 + \cdots + r_m(\cdot)s_m$. Therefore, by Lemma 1,

$$r_j(P) > r_k(P) \Longrightarrow s_j \geq s_k \tag{1.6}$$

for each $j, k = 1, 2, \ldots, m$. Since $r_j(P) > r_k(P)$ if and only if $a_j \; P \; a_k$, we have

$$a_j \; P \; a_k \Longrightarrow s_j \geq s_k \; \forall a_j, a_k \in A.$$

Therefore, $P \in F^{LB}(\succsim)$, which in turn implies $F^{LC}(\succsim) \subset F^{LB}(\succsim)$.

Next, we show that $F^{LB}(\succsim) \subset F^{LC}(\succsim)$. Take any $P \in F^{LB}(\succsim)$. By the definition of r, $a_j \; P \; a_k$ if and only if $r_j(P) > r_k(P)$. Moreover, by the definition of the linear Borda rule (1.4), $a_j \; P \; a_k$ implies $s_j \geq s_k$ for each $j, k = 1, 2, \ldots, m$. These together imply that

$$r_j(P) > r_k(P) \Longrightarrow s_j \geq s_k \tag{1.7}$$

for each $j, k = 1, 2, \ldots, m$. By (1.7) and Lemma 1, P maximizes $r_1(\cdot)s_1 + r_2(\cdot)s_2 + \cdots + r_m(\cdot)s_m$. By (1.5), P is a maximizer of $\sum_{i \in I} C(r(\cdot), r(\succsim_i))$. Hence, $P \in F^{LC}(\succsim)$, that is, $F^{LB}(\succsim) \subset F^{LC}(\succsim)$. $\square$

1.4.2 Cosine similarity and scoring rules

In Section 2, we employed rank expression r to convert preferences to numerals. The rank expression of $\succsim_i \in \mathcal{P}$ is

$$r(\succsim_i) = (r_{a_1}(\succsim_i), r_{a_2}(\succsim_i), \ldots, r_{a_m}(\succsim_i)) \in \mathbb{N}^m,$$

where

$$r_{a_k}(\succsim_i) = |\{a' \in A : a_k \succsim_i a'\}|.$$

Consider other ways to convert preferences to numerals. Consider any vector $s = (s_1, s_2, \ldots, s_m) \in \mathbb{R}^m_{++}$ with $s_1 < s_2 < \cdots < s_m$. For each $\succsim_i \in \mathcal{P}$ and each $a_j \in A$, we define $s_{a_j}(\succsim_i) \in \mathbb{R}_{++}$ be such that

$$s_{a_j}(\succsim_i) = s_k,$$

where $k = r_{a_j}(\succsim_i)$. Then we obtain a vector

$$s(\succsim_i) = (s_{a_1}(\succsim_i), s_{a_2}(\succsim_i), \ldots, s_{a_m}(\succsim_i)) \in \mathbb{R}^m_{++}$$

that expresses $\succsim_i$. We call $s(\succsim_i)$ s-expression of $\succsim_i$. We show that if we employ s-expression, the cosine similarity rule is equivalent to the scoring rule associated with the score vector s. The scoring rule with s, denoting F^s, is the ranking rule such that

$$a_j \ F^s(\succsim) \ a_k \text{ if and only if } \sum_{i \in N} s_{a_j}(\succsim_i) \geq \sum_{i \in N} s_{a_k}(\succsim_i)$$

for each $\succsim \in \mathcal{P}^n$ and $a_j, a_k \in A$. The Borda rule is the scoring rule associated with $s = (1, 2, 3, \ldots, m)$. Therefore, Theorem 1 is a special case of the following result. Since its proof parallels that of Theorem 1, we state the result as a corollary.

Corollary 1. *Let $s = (s_1, s_2, \ldots, s_m) \in \mathbb{R}^m_{++}$ with $s_1 < s_2 < \cdots < s_m$. If we use s-expression, the cosine similarity rule is equivalent to the scoring rule associated with s.*

1.5 Conclusion

We have proposed a voting rule that chooses a social ranking maximizing the cosine similarity, namely, the cosine similarity rule. We have shown that the cosine similarity rule coincides with the Borda rule. Cosine similarity appears often in the literature of computer science to measure similarity, so our analysis provides a rationale for the use of the Borda rule based on cosine similarity. Moreover, we have discussed an analogous relation between Borda's rule and Condorcet's rule. The choice between these two rules can be attributed to the choice between the Kemeny minimization and the cosine similarity maximization.

Cosine similarity clearly satisfies some standard properties such as symmetry, scale invariance, and neutrality but violates axioms of distance and some independence axioms such as local independence. One of important future research is to characterize the cosine similarity measure. If we can characterize cosine similarity by some axioms and can compare the axioms with those of Kemeny distances, the analysis helps us answer the two-century old question “Condorcet or Borda?"

Appendix 1

Proof of Lemma 1. Consider any $x, y \in \mathbb{R}^m$ with $x_1 \leq x_2 \leq \cdots \leq x_m$ and $y_1 \leq y_2 \leq \cdots \leq y_m$. Take any permutation σ on $\{1, 2, \ldots, m\}$.

First, we shall show that if $\sigma(m) \neq m$, then there exists a permutation π on $\{1, 2, \ldots, m\}$ such that

$$\sum_{k=1}^{m} x_k y_{\pi(k)} \geq \sum_{k=1}^{m} x_k y_{\sigma(k)}. \tag{1.8}$$

Assume that $\sigma(m) \neq m$. Then, there exist $i, j < m$ such that $\sigma(m) = i$ and $\sigma(j) = m$. Since $x_m \geq x_j$ and $y_m \geq y_i$, $(x_m - x_j)(y_m - y_i) \geq 0$. Therefore,

$$x_m y_m + x_j y_i \geq x_m y_i + x_j y_m. \tag{1.9}$$

Let π be the permutation on $\{1, 2, \ldots, m\}$ such that

$$\pi(k) = \begin{cases} m & \text{if} \quad k = m, \\ i & \text{if} \quad k = j, \\ \sigma(k) & \text{if} \quad k \notin \{m, j\}. \end{cases}$$

By (1.9),

$$\begin{aligned} \sum_{k=1}^{m} x_k y_{\pi(k)} &= \sum_{k \neq m, j} x_k y_{\sigma(k)} + x_m y_m + x_j y_i \\ &\geq \sum_{k \neq m, j} x_k y_{\sigma(k)} + x_m y_i + x_j y_m \\ &\geq \sum_{k=1}^{m} x_k y_{\sigma(k)}, \end{aligned}$$

that is, we have (1.8).

Next, we show that if $\sigma(m) = m$ and $\sigma(m-1) \neq m - 1$, then there exists a permutation π

such that

$$\sum_{k=1}^{m} x_k y_{\pi(k)} \geq \sum_{k=1}^{m} x_k y_{\sigma(k)}. \tag{1.10}$$

Assume that $\sigma(m) = m$ and $\sigma(m-1) \neq m-1$. Then, there exist $i, j < m-1$ such that $\sigma(m-1) = i$ and $\sigma(j) = m-1$. Since $x_{m-1} \geq x_j$ and $y_{m-1} \geq y_i$, we have $(x_{m-1} - x_j)(y_{m-1} - y_i) \geq 0$. Therefore,

$$x_{m-1}y_{m-1} + x_j y_i \geq x_{m-1}y_i + x_j y_{m-1}. \tag{1.11}$$

Let π be the permutation on $\{1, 2, \ldots, m\}$ such that

$$\pi(k) = \begin{cases} m-1 & \text{if} \quad k = m-1, \\ i & \text{if} \quad k = j, \\ \sigma(k) & \text{if} \quad k \notin \{m-1, j\}. \end{cases}$$

By (1.11),

$$\begin{aligned} \sum_{k=1}^{m} x_k y_{\pi(k)} &= \sum_{k \neq m-1, j} x_k y_{\sigma(k)} + x_{m-1}y_{m-1} + x_j y_i \\ &\geq \sum_{k \neq m-1, j} x_k y_{\sigma(k)} + x_{m-1}y_i + x_j y_{m-1} \\ &\geq \sum_{k=1}^{m} x_k y_{\sigma(k)}, \end{aligned}$$

that is, we have (1.10).

In a general case, take any $\ell \in \{1, 2, \ldots, m\}$ and assume that $\sigma(m) = m, \sigma(m-1) = m-1, \ldots, \sigma(\ell+1) = \ell+1$, and $\sigma(\ell) \neq \ell$. Let π be identical to σ except that $\pi(\ell) = \ell$ and $\pi(\sigma^{-1}(\ell)) = \sigma(\ell)$. Then,

$$\sum_{k=1}^{m} x_k y_{\pi(k)} \geq \sum_{k=1}^{m} x_k y_{\sigma(k)}$$

by the similar argument. Therefore, $\sum_{k=1}^{m} x_k y_k \geq \sum_{k=1}^{m} x_k y_{\sigma(k)}$. $\square$

Chapter 2

Consistent Representation and Electoral Systems

2.1 Introduction

Electoral systems are widely used to select governmental representatives in democratic countries. Since representatives usually belong to political parties, one important function of electoral systems is to determine the distribution of seats among political parties in the congress. However, it is a controversial proposition that the distribution of seats in the congress can appropriately reflect preferences of the people in the nation. The following example illustrates a paradoxical situation wherein the distribution of seats does not appropriately reflect preferences of the people. Consider the example of a nation containing 100 voters who have preferences on political parties X, Y, and Z. They elect five representatives using a single-member district system with five districts $d_1, d_2, \ldots, d_5$. There are 20 voters in each district. Consider the following preference profile.

d_1			
5	9	3	3
X	Z	Z	X
Z	X	Y	Y
Y	Y	X	Z

d_2		
7	11	2
X	Z	Y
Z	X	Z
Y	Y	X

d_3			
8	5	3	4
Y	Y	X	Z
X	Z	Y	X
Z	X	Z	Y

d_4		
11	5	4
Y	X	Z
X	Y	Y
Z	Z	X

d_5			
5	6	4	5
Y	Y	X	Z
X	Z	Z	Y
Z	X	Y	X

Table 1: Preference profile

For example, in district d_1, five voters prefer X to Z and Z to Y, and in district d_2, eleven voters prefer Z to X and X to Y. Now, the total number of seats to fill in the congress is five. This scenario illustrates the questions: Which political party should get a large number of seats in the congress? What division would make the legislative chamber truly representative?

To tackle these questions, we focus on two central notions in social choice theory: Condorcet winners and Borda winners. Most voting rules are designed to choose a Condorcet winner, a Borda winner or a near alternative, and it is known that such voting rules have many desirable properties.[*8] Thus, we aim here to require that a political party that is a Condorcet winner or a Borda winner should receive a large number of seats in the congress, and a political party that is a Condorcet loser or a Borda loser should receive few number of seats in the congress.[*9] In this preference profile (Table 1),

- Party X is the Condorcet winner and the Borda winner in the nation;
- Party Y is the Condorcet loser and the Borda loser in the nation.[*10]

Therefore, we expect that X receives more seats in the congress than Y. However, if we employ the plurality rule in each district, Y gets three seats (in $d_3, d_4,$, and d_5), Z gets two seats (in d_1 and d_2), and X gets no seat in the congress. Moreover, even if scoring rules, particularly the Borda rule, or voting rules satisfying Condorcet's principle are employed, the same result occurs. The house of representatives is dominated by the Condorcet loser and Borda loser even if desirable voting rules are applied. We call this peculiar situation the *Loser Dominance Paradox*. Figure 2 illustrates this paradox.

The Loser Dominance Paradox. In this preference profile (Table 1), Y is the Condorcet and Borda loser in the nation but it is the dominant party (gains the absolute majority number

[*8] Black's method always chooses a Condorcet winner or a Borda winner (Black 1958). The Kemeny-Young-Method and the Copeland method choose a Condorcet winner or one of the nearest alternatives, and these rules have desirable properties (Kemeny 1959, Young and Levenglick 1978). Young (1974, 1975) characterize the Borda rule and scoring rules by sets of desirable properties.

[*9] A Condorcet *winner* is an alternative that beats any other alternative in pairwise-majority comparison. Conversely, a Condorcet *loser* is an alternative beaten by any other alternative in pairwise-majority comparison. A Borda *winner* is an alternative that gets the highest Borda score, and a Borda *loser* is one that gets the least Borda score. Formal definitions are given in Section 2.2.

[*10] Let us show these assertions. X is the Condorcet winner and Y is the Condorcet loser because X beats Y and Z, and Y is beaten by Z in pairwise-majority comparisons. Indeed, 51 (= 17 + 18 + 7 + 5 + 4) voters prefer X to Y, 51 (= 8 + 7 + 11 + 16 + 9) voters prefer X to Z and 52 (= 17 + 18 + 4 + 4 + 9) voters prefer Z to Y.

X is the Borda winner in the nation and Y is the Borda loser in the nation because the Borda score of X is 202 $(= 3 \cdot 27 + 2 \cdot 48 + 1 \cdot 25)$, that of Y is 197 $(= 3 \cdot 37 + 2 \cdot 23 + 1 \cdot 40)$, and that of Z is 201 $(= 3 \cdot 36 + 2 \cdot 29 + 1 \cdot 35)$.

of seats) in the congress if some desirable voting rules are used in each district. Moreover, X is the Condorcet and Borda winner in the nation but it gets no seats if some desirable voting rules are used in each district. More precisely,

(i) if we use the plurality rule in each district, Y is the dominant party and X gets no seats;

(ii) if we use any voting rule satisfying Condorcet's principle, Y is the dominant party and X gets no seats;

(iii) if we use any scoring rule (including the Borda rule) associated with a score vector (a, b, c) with $a > b \geq c$, Y is the dominant party and X gets no seats.

Proof. See Appendix 2. $\square$

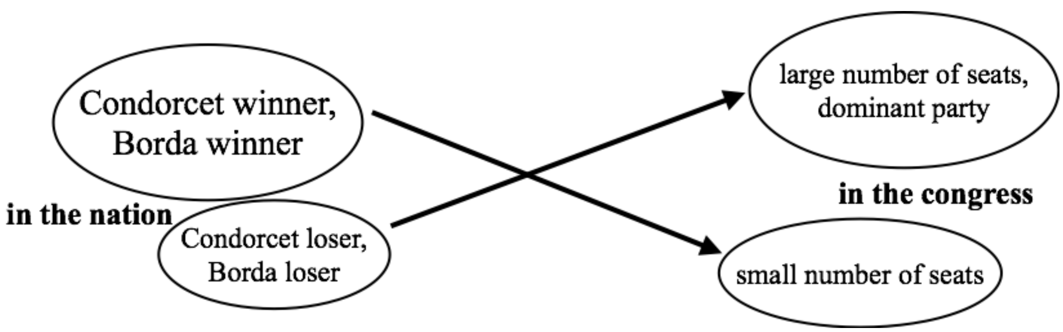

Figure 2: The Loser Dominance Paradox

This argument strongly suggests that a single-member district system may not appropriately reflect preferences of the people. We define *Condorcet consistency* and *Borda consistency* to analyze electoral systems that can avoid such a paradoxical situation. Condorcet (Borda) consistency requires that whenever a Condorcet (Borda) winner and a Condorcet (Borda) loser exist in the nation, the number of seats in the congress gained by the winner be greater than or equal to that gained by the loser. Our research takes the same approach as Borda (1784), who pointed out that the plurality rule may choose a Condorcet loser and searched for alternative voting rules to avoid choosing a Condorcet loser. Our purpose is to point out a fatal defect in some prevalent electoral systems and to search for alternative electoral systems to avoid dominance by Condorcet and Borda losers.

Situations similar to the Loser Dominance Paradox can emerge in real life. A recent prominent example is the 2016 United States president election. In the election, Donald Trump obtained the required majority of votes of *state electors* and defeated Hillary Clinton. However, the number of *nationwide popular votes* for Trump was less than that cast for Clinton. In democratic countries, it is commonly observed that voters cast a ballot to choose electors or representatives at the first stage, and the electors or representatives vote on important issues at the second stage, however, such two-stage of electoral systems are vulnerable to this Loser Dominance Paradox.

The Ostrogorski Paradox (Rae and Daudt 1976) and Simpson's Paradox (Good and Mittal 1987) are two well-known paradoxes that exhibit conflicts between direct aggregation and indirect aggregation. Recent theoretical studies discussing consistent representative democracy have been conducted by Chambers (2008, 2009). He analyzes a certain class of hierarchial voting procedures that are robust against gerrymandering, but these studies are independent from our study.

Our first main result is an impossibility theorem on single-member district systems, which is a generalization of the Loser Dominance Paradox. The theorem states that in a single-member district system, under weak conditions, if the plurality rule, Condorcet consistent rule, or Borda rule are employed, the electoral system is neither Condorcet consistent nor Borda consistent (Theorem 1). Next, we focus on proportional representation systems. We define a wide class of proportional representation systems including simple proportional representation systems. We show that a simple proportional representation system is neither Condorcet consistent nor Borda consistent (Proposition 1). Finally, we propose a new electoral system, namely, the *Borda proportional representation system*. We show that a Borda proportional representation system is both Condorcet consistent and Borda consistent (Theorem 2).

The rest of this chapter is organized as follows. In Section 2.2, we introduce our model and definitions. In Section 2.3, we define single-member district systems and show impossibility results. In Section 2.4, we define proportional representation systems and discuss desirability of Borda proportional representation systems. Section 2.5 concludes this chapter.

2.2 Definitions

Let $N = \{1, 2, \dots, n\}$ be the set of voters with $n \geq 3$ and $P = \{p_1, p_2, \dots, p_m\}$ be the set of political parties with $m \geq 3$. Let $\mathcal{R}$ be the set of orderings on P.[*11] Each voter $i \in N$ has a preference relation $\succsim_i \in \mathcal{R}$ on the set of political parties P. [*12] A *preference profile* is a list of preferences

$$\succsim = (\succsim_1, \succsim_2, \dots, \succsim_n) \in \mathcal{R}_N.$$

For each $N' \subset N$, $\succsim_{N'} = (\succsim_i)_{i \in N'} \in \mathcal{R}_{N'}$ denotes the preference profile of group N'. Let a natural number $S \in \mathbb{N}$ be the number of seats to fill. Let $X \subset \mathbb{Z}_+^m$ be the set of possible distributions of seats in the congress, that is,

$$X = \left\{ (S_1, S_2, \dots, S_m) \in \mathbb{Z}_+^m : \sum_{p_j \in P} S_j = S \right\}.$$

2.2.1 Electoral systems

An *electoral system* $E : \mathcal{R}_N \to X$ maps each preference profile to a distribution of seats in the congress. We also write an electoral system $E = (D, A, F)$ because it consists of three components: a set of *electoral districts*, *apportionments*, and *voting rules*. A set of *electoral districts* is $D = \{d_1, d_2, \dots, d_t\}$ with $t \geq 1$ where $d_1, d_2, \dots, d_t$ are partitions of N, that is, $\cup_{d_k \in D} d_k = N$ and for each two districts $d_k, d_\ell \in D$, $d_k \cap d_\ell = \emptyset$. An *apportionment* is a function $A : D \to \mathbb{Z}_+$ with

$$\sum_{d_k \in D} A(d_k) = S \text{ and } A(d_k) \geq 1 \text{ for each } d_k \in D.$$

For instance, $A(d_1) = 3$ indicates that the number of apportioned seats or the number of winners of the district d_1 is 3. If E is a single-member district electoral system, $A(d_k) = 1$

[*11] A binary relation $\succsim_i$ is an ordering if it is *complete* and *transitive*. A binary relation $\succsim_i$ is *complete* if for any $a, b \in P$, $a \succsim_i b$ or $b \succsim_i a$ holds. It is *transitive* if for any $a, b, c \in P$, $[a \succsim_i b$ and $b \succsim_i c]$ implies $a \succsim_i c$.

[*12] In reality, each political party puts up one candidate for each district and each voter has corresponding preferences on the set of candidates.

for each $k = 1, \ldots, t$. A *voting rule* is a function $F : \cup_{d_k \in D} \mathcal{R}_{d_k} \to \mathbb{Z}^m_+$ with

$$\sum_{\ell=1}^{m} F_\ell(\succsim_{d_k}) = A(d_k) \text{ for each } d_k \in D,$$

where $F_\ell(\succsim_{d_k})$ is the ℓ-th component of $F(\succsim_{d_k})$. For example, $F(\succsim_{d_1}) = (1, 0, 2, 0)$ means that the party p_1 wins 1 seat, the party p_3 wins 2 seats, and the party p_2 and p_4 win no seat in the district d_1. Let $\mathcal{E}$ be the set of electral systems $E = (D, A, F)$.

For each $\succsim \in \mathcal{R}_N$, each $E = (D, A, F) \in \mathcal{E}$, and each party $p_\ell \in P$, let us denote the number of winning seats of p_ℓ by $S_\ell(\succsim, E) \in \mathbb{Z}_+$, that is,

$$S_\ell(\succsim, E) = \sum_{d_k \in D} F_\ell(\succsim_{d_k}).$$

When the people's preference profile is $\succsim$, the distribution of seats in the congress under the electoral system E is

$$S(\succsim, E) = (S_1(\succsim, E), S_2(\succsim, E), \ldots, S_m(\succsim, E)) \in X.$$

2.2.2 Consistency conditions

Before defining consistency conditions, we offer some definitions that play important rolls in our analysis. We introduce Condorcet winners and Borda winners as *desirable* alternatives that should prevail. On the other hand, we introduce Condorcet losers and Borda losers as *undesirable* alternatives that should not prevail.

A *Condorcet winner* is a candidate that defeats every other candidate in pairwise-majority comparisons. That is, a party $p_\ell \in P$ is a *Condorcet winner* in $\succsim \in \mathcal{R}_N$ if

$$\begin{aligned} &|\{i \in N : p_\ell \succsim_i p_j\}| \geq \tfrac{1}{2}n \quad \text{for all } p_j \in P, \text{ and} \\ &|\{i \in N : p_\ell \succsim_i p_k\}| > \tfrac{1}{2}n \quad \text{for som } p_k \in P. \end{aligned}$$

A party $p_\ell \in P$ is a *Condorcet loser* in $\succsim \in \mathcal{R}_N$ if

$$|\{i \in N : p_\ell \succsim_i p_j\}| \leq \tfrac{1}{2}n \quad \text{for all } p_j \in P, \text{ and}$$

$$|\{i \in N : p_\ell \succsim_i p_k\}| < \tfrac{1}{2}n \quad \text{for som } p_k \in P.$$

For each party $p_\ell \in P$ and each preference $\succsim_i \in \mathcal{R}$, let

$$r_\ell(\succsim_i) \equiv \frac{|\{p_j \in P : p_\ell \succsim p_j\}| + |\{p_j \in P : p_\ell \succ p_j\}| + 1}{2}$$

be the *inverse ranking* of $p_\ell \in P$ at $\succsim_i$. For example, $r_\ell(\succsim_i) = m$ means that the party p_ℓ is most preferred by voter i, and $r_\ell(\succsim_i) = 1$ means that the party p_ℓ is the worst alternative for voter i. Note that if $\{p_j \in P : p_\ell \sim p_j\} = \{p_\ell\}$, $r_\ell(\succsim_i) = |\{p_j \in P : p_\ell \succsim p_j\}|$. We call $B_\ell(\succsim) \equiv \sum_{i \in N} r_{p_\ell}(\succsim_i)$ the *Borda score* of $p_\ell \in P$ in $\succsim$. A *Borda winner* is a candidate that gets the highest Borda score. That is, a party $p_\ell \in P$ is a *Borda winner* in $\succsim \in \mathcal{R}_N$ if

$$B_\ell(\succsim) \geq B_j(\succsim) \text{ for all } p_j \in P \text{ and } B_\ell(\succsim) > B_k(\succsim) \text{ for some } p_k \in P.$$

A *Borda loser* is a candidate that gets the least Borda score. That is, a party $p_\ell \in P$ is a *Borda loser* in $\succsim \in \mathcal{R}_N$ if

$$B_\ell(\succsim) \leq B_j(\succsim) \text{ for all } p_j \in P \text{ and } B_\ell(\succsim) < B_k(\succsim) \text{ for some } p_k \in P.$$

We require that if a party is a Condorcet winner or a Borda winner, and another party is a Condorcet loser or a Borda loser in the nation, then the number of seats of the winner should be greater than that of the loser in the congress.

Condorcet consistency. An electoral system $E \in \mathcal{E}$ is *Condorcet consistent* if for every $\succsim \in \mathcal{R}_N$, if there exists p_ℓ and $p_j \in P$ such that p_ℓ is a Condorcet winner in $\succsim$ and p_j is a Condorcet loser in $\succsim$, then

$$S_\ell(\succsim, E) \geq S_j(\succsim, E).$$

This requirement is an extension of Condorcet's principle from social choice functions to electoral systems. In a similar way, we define Borda consistency for electoral systems.

Borda consistentency. An electoral system $E \in \mathcal{E}$ is *Borda consistent* if for every $\succsim \in \mathcal{R}_N$, if $p_\ell \in P$ is a Borda winner in $\succsim$ and p_j is a Borda loser, then

$$S_\ell(\succsim, E) \geq S_j(\succsim, E).$$

If an electoral system violates these conditions, the Loser Dominance Paradox occurs. We search for electoral systems satisfying these conditions.

2.3 Single-member district electoral systems

A single-member district system is one of the commonly used electoral systems. We introduce its formal definition.

Single-member district electoral system. We call $E = (D, A, F) \in \mathcal{E}$ a *single-member district system* if for each $d_k \in D$, $A(d_k) = 1$. Clearly, $|D| = S$. Let $\mathcal{E}^s \subset \mathcal{E}$ be the set of single-member district systems.

Since the number of winners in each district is one, we can use familiar voting rules such as the plurality rule as voting rules F. We focus on three central voting rules; Condorcet rules, the Borda rule, and the plurality rule.

Condorcet rule. In a single-member district system $E = (D, A, F) \in \mathcal{E}^s$, a voting rule F is *Condorcet* if for each $\succsim_{d_k} \in \cup_{d_k \in D} \mathcal{R}_{d_k}$, whenever a *Condorcet winner* in $\succsim_{d_k}$ exists,[*13]

$$F_\ell(\succsim_{d_k}) = 1 \Longrightarrow p_\ell \text{ is a \textit{Condorcet winner} in } \succsim_{d_k}.$$

Borda rule. In a single-member district system $E = (D, A, F) \in \mathcal{E}^s$, a voting rule F is a *Borda* rule if for each $\succsim_{d_k} \in \cup_{d_k \in D} \mathcal{R}_{d_k}$,

$$F_\ell(\succsim_{d_k}) = 1 \Longrightarrow p_\ell \in \arg\max_{p_j \in P} B_j(\succsim_{d_k}).$$

Plurality rule. In a single-member district system $E = (D, A, F) \in \mathcal{E}^s$, a voting rule F is a

[*13] A political party $p_\ell \in P$ is a *Condorcet winner* in $\succsim_{d_k}$ if $|\{i \in d_k : p_\ell \succsim_i p_j\}| > \frac{1}{2}|d_k|$ for all $p_j \in P$.

plurality rule if for each $\succsim_{d_k} \in \cup_{d_k \in D} \mathcal{R}_{d_k}$,

$$F_\ell(\succsim_{d_k}) = 1 \Longrightarrow p_\ell \in \arg\max_{p_j \in P} |\{i \in d_k : i \text{ prefers } p_j \text{ best}\}|.$$

We first show that any single-member district system with Condorcet voting rules, Borda rules, or plurality rules is neither Condorcet consistent nor Borda consistent under some weak conditions. Although we employ desirable voting rules in each district, the Condorcet winner in the nation may be the weakest party in the congress. Moreover, it can not avoid that the Condorcet loser in the nation becomes the dominant party in the congress. This suggests incompatibility between desirable voting rules and a minimum requirement for electoral systems under single-member district systems.

Theorem 1. *Suppose the number of seats in the congress $S \geq 9$.*[*14]*. Let $E = (D, A, F)$ be a single-member district electoral system.*

(i) If $|d_k| \geq 3$ for each $d_k \in D$, E is not Condorcet consistent for any Condorcet consistent voting rule, the Borda rule, or the plurality rule F.

(ii) If $|d_k| \geq 8$ for each $d_k \in D$, E is not Borda consistent for any Condorcet consistent voting rule, the Borda rule, or the plurality rule F.

Proof. See Appendix 2. □

In the proof of Theorem 1, we have shown that in any single-member district electoral system, the weakest political party (both the Borda loser and the Condorcet loser) may be the dominant party in the congress if voting rules are Condorcet, Borda or the plurality rule.

Corollary 1. *Suppose $S \geq 9$ and let $E = (D, A, F) \in \mathcal{E}^s$ be a single-member district electoral system with $|d_k| \geq 8$ for each $d_k \in D$. If F is Condorcet, Borda, or the plurality rule, there exists a preference profile $\succsim \in \mathcal{R}_N$ such that there exists a political party $p_\ell \in P$ such that p_ℓ is both the Condorcet loser and the Borda loser in $\succsim$ but it is the dominant party in the congress.*

[*14]In fact, when $S = 5$ or 7, (i) holds but when $S = 6$ or 8, (i) does not hold. Although $S \geq 9$ is not a necessary condition and the lower bound of S for (i) is 5, to simplify the statement of this theorem, we wrote $S \geq 9$.

2.4 Proportional representation systems

In the previous section, we confronted with impossibility of consistent representation under single-member district electoral systems. We next consider whether a proportional representation system satisfies Condorcet or Borda consistency. There are many types of proportional representation systems such as Jefferson (d'Hondt), Webster, and the quota method because the seats of the congress are indivisible and there are different ways to deal with fractions. [*15]

Unfortunately, we show that a proportional representation system is neither Condorcet consistent nor Borda consistent for any reasonable way to round fractions.

Proposition 1. *A proportional representation system is neither Condorcet consistent nor Borda consistent.*

Proof. (Proof by example). Let $n = 100$, $P = \{x, y, z\}$, and $S = 30$. Define $\succsim \in \mathcal{R}_N$ as follows (Table 2):

30	10	30	30
x	x	y	z
y	z	z	y
z	y	x	x

Table 2: PR is neither Condorcet nor Borda consistent.

In a proportional representation system, seats of (x, y, z) are distributed proportionally to $40 : 30 : 30$, so the distribution of seats is

$$(S_x, S_y, S_z) = (12, 9, 9).$$

So, the political party x gets the most seats in the congress. However, x is the Condorcet loser and the Borda loser, and y is the Condorcet winner and the Borda winner.[*16] Therefore, a proportional representation system violates both of Condorcet consistency and Borda consistency. □

[*15] Balinski and Young (1975, 1978 and 1982) conduct axiomatic analysis on federal apportionment methods and proportional representation systems.

[*16] In majority comparisons, $x : y = 40 : 60$, $x : z = 40 : 60$, $y : z = 60 : 40$, so x is the Condorcet loser and y is the Condorcet winner. $B_x = 180, B_y = 220. B_z = 200$, so x is the Borda loser and y is the Borda winner.

Single-member electoral systems and proportional representation systems are two of the most prevalent electoral systems, but they are neither Condorcet nor Borda consistent. Under these electoral systems, the Loser Dominance Paradox may occur. To avoid the paradox, we introduce a new electoral system that satisfies both consistency conditions, namely, the *Borda proportional representation system*. To define it formally, we introduce the definition of *generalized proportional representation systems*. First, we define a *rounding function* $[\cdot]_r$ that maps each real number to an integer. A function $[\cdot]_r$ is a rounding function if it is nondecreasing and for each $x \in \mathbb{R}_+$, $[x]_r \in \mathbb{Z}_+$ and

$$x - 1 \leq [x]_r \leq x + 1.$$

For example, the floor function $[\cdot]$ is a rounding function and $[x]_r = [x + \frac{1}{2}]$ is also a rounding function. All of our results hold for any rounding function. Next, we define a *score function* $g : \mathcal{R}^N \to \mathbb{R}_+^m$ that maps each preference profile to an m-dimensional integer vector whose component $g_\ell(\succsim)$ means the *score* of p_ℓ. In a generalized proportional representation system, seats are distributed proportionally to scores of parties.

Generalized proportional representation system. We call $E = (D, A, F) \in \mathcal{E}$ a *generalized proportional representation system* if $D = \{N\}$, $A(N) = S$, and F is *proportional*. A voting rule $F : \mathcal{R}^N \to \mathbb{Z}_+^m$ is *proportional* if there exists a rounding function $[\cdot]_r$ and a score function g such that for each $\succsim \in \mathcal{R}^N$, there exists $\lambda \in \mathbb{R}_{++}$ such that for each $\ell = 1, 2, \ldots, m$,

$$F_\ell(\succsim) = \left[\frac{g_\ell(\succsim)}{\lambda}\right]_r,$$

and

$$\sum_{p_\ell \in P} F_\ell(\succsim) = S.^{*17}$$

Let $\mathcal{E}^p \subset \mathcal{E}$ be the set of generalized proportional representation systems.

A proportional representation system is one of the generalized proportional representation

[*17] There may be tie-cases. Tie-breaking should be done monotonically to g_ℓ. For example, consider the case that $S = 1$, $g_1 = g_2 = 10$, and $g_3 = 1$. Then any $\lambda > 0$ and $[\cdot]_r$ cannot satisfy $F_1 + F_2 + F_3 = S$, so we need a tie-breaking. In this case, monotonic tie-breaking rule results in either one of F_1 or F_2 equals to one and $F_3 = 0$ because $g_1 = g_2 > g_3$. Our all results hold as long as a tie-breaking rule is monotonically to g_ℓ.

systems, which is based on the *simple score function*. The *simple score function* g^s is a score function such that

$$g^s_\ell(\succsim) = |\{i \in N : i \text{ prefers } p_\ell \text{ best at } \succsim_i\}|$$

for each $p_\ell \in P$. We can consider various proportional representation systems choosing score functions. A principal example of them is a *Borda proportional representation system*, which is based on the *Borda score function* g^B. It is a score function such that $g^B_\ell(\succsim) = B_\ell(\succsim)$, the Borda score of p_ℓ, for each party $p_\ell \in P$. We call a generalized proportional representation system with g^B a *Borda proportional representation system*. Under the system, seats are distributed proportionally to Borda scores that parties gain. Consider the same preference profile considered in the proof of Proposition 1. Now y is the Borda and Condorcet winner, and x is the Borda and Condorcet loser. In a Borda proportional representation system, seats are distributed proportionally to $B_x : B_y : B_z = 180 : 220 : 200$, so the distribution of seats is

$$(S_x, S_y, S_z) = (9, 11, 10).$$

So, the number of seats of winners is larger than that of losers. This holds in general.

Theorem 2. *A Borda proportional representation system is Condorcet consistent and Borda consistent.*

Proof. See Appendix 2. □

A Borda proportional representation system is desirable because it satisfies both consistency conditions, but it has a shortcoming. It may be vulnerable to existence of minor candidates or candidates unworthy for serious consideration. That is, under a Borda proportional representation system, a very minor political party can often get not a few number of seats in the congress. For instance, let $S = 40$ and consider the following profile. In

30	10	30	30
y	z	y	z
z	y	z	y
x	x	x	x

this profile, party x is the worst alternative for all voters. So, it may be a “bad” candidate or unworthy for serious consideration. However, under a Borda proprtional representation system, since $B_x : B_y : B_z = g_x : g_y : g_z = 100 : 260 : 240$, we have $S_x = 7, S_y = 17$ and $S_z = 16$. Even if all voters did not want to give party x any seat in the congress, x can get 7 seats, which occupies not a little portion of the total. In addition, since no party is dominant in the congress, x may be a pivotal.

We can solve this problem by transforming score functions monotonically. For example, let $h(s) = s^3$ and $g_\ell(\succsim) = h(B_\ell(\succsim))$ for each p_ℓ. Then since $g_x : g_y : g_z = 100^3 : 260^3 : 240^3$, we have $S_x = 1, S_y = 22$ and $S_z = 17$. Considering another example, let

$$h^Q(s) = \begin{cases} s & \text{if } s \geq Q \\ 0 & \text{otherwise} \end{cases}$$

for some quota $Q \geq 0$. The quota Q can depend on m and n. For example, Q can be the average Borda score, that is,

$$\frac{\sum_{p_\ell \in P} B_\ell}{m} = \frac{1}{2} n(m+1).$$

In this case, $Q = \frac{1}{2} \cdot 100 \cdot 4 = 200$. Then since $g_x : g_y : g_z = 0 : 260 : 240$, $S_y = [20.8]_r$, $S_z = [19.2]_r$, and $S_x = 0$. In general, let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing function and g^{hB} be the score function defined as $g^{hB}_\ell = h(g^B_\ell)$ for each $\ell = 1, 2, \ldots, m$. Then, a proportional representation system with g^{hB} satisfies Condorcet and Borda consistency, and by choosing a suitable h, we can avoid the problem of very minor candidates. So, there are many types of proportional representation systems satisfying Borda consistency and Condorcet consistency. However, a proportional representation system based on positional rules other than the Borda rule cannot satisfy consistency conditions. That is because the Borda rule is the unique positional rule that always ranks the Condorcet winner higher than the Condorcet loser (Saari 1990).

We formalize this argument. A score function $g : \mathcal{R} \to \mathbb{R}^m$ is a *proper* score function if there exists a vector $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$ with $a_1 \leq \cdots \leq a_m$, for any $\succsim \in \mathcal{R}^n$ and any $\ell \in P$,

$$g_\ell(\succsim) = \sum_{i \in N} a_{r_\ell(\succsim_i)},$$

where $r_\ell(\succsim_i)$ denotes the inverse ranking of p_ℓ at $\succsim_i$. In other words, a score function is proper if its score is generated by some *positional rule*. The simple majority rule (plurality rule) and the Borda rule are positional rules, so the simple score function and the Borda score functions are proper. An axiomatization of positional rules is established by Young (1975), so we here do not devote us to consider desirability of proportional representation systems with proper score functions. A generalized proportional representation system $E \in \mathcal{E}^p$ is *based on* a score function ϕ if there exists a non-decreasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $\succsim \in \mathcal{R}^n$ and any $\ell \in P$,

$$g_\ell(\succsim) = h(\phi_\ell(\succsim)).$$

Then we have the following corollary.

Corollary 2. *Let $E \in \mathcal{E}^p$ be a proportional representation system based on a proper score function. E is Borda and Condorcet consistent if and only if it is based on the Borda score function.*

2.5 Conclusion

We have provided the Loser Dominance Paradox in single-member district systems. It displays that for any Condorcet consistent voting rule or scoring rules, the Condorcet and Borda winner party may not get any seat in the congress and the Condorcet and Borda loser party gets an over-half number of seats in the congress. We define the Condorcet consistency and the Borda consistency to analyze electoral systems that can avoid this paradox. First, we have shown the incompatibility of these consistency conditions and desirable voting rules in single-member district systems. Moreover, we have shown that simple proportional representation systems violates these consistency conditions. Finally, we have proposed the Borda proportional representation system and showed that it satisfies both consistency conditions.

In our analysis, we have focused on score functions defined by positional rules. Note that there exists a generalized proportional representation system that is Condorcet and Borda consistent and not based on proper score functions. A prominent example is a *Kemeny proportional representation system*. Let $K_\ell(\succsim)$ be the Kemeny ranking of p_ℓ and $g^K_\ell(\succsim) = m - K_\ell(\succsim)$. Then a generalized proportional representation system with g^K, namely the *Kemeny proportional representation system*, satisfies Condorcet and Borda consistency.[*18]

[*18]Of course, we can transform the score function g^K into any other one as long as the order of scores is

That is because the Kemeny rule ranks a Condorcet winner at the top rank if it exists (Young and Levenglick 1978). Moreover, in Kemeny rankings, Borda winners are always ranked higher than Borda losers (Saari and Merlin 2000). In summary, proportional representation systems based on the Borda score function g^B or the Kemeny score function g^K satisfy both consistency conditions. Characterizing a class of electoral systems by Condorcet and Borda consistency and other desirable properties is an important future research.

Appendix 2

Proof of the Loser Dominance Paradox

Proof. (i) and (ii) can be shown easily. We only show (iii). Let F be any scoring rule associated with the score vector (a, b, c) with $a > b \geq c$. Let $s(p, d)$ be the score of party $p \in \{X, Y, Z\}$ in district $d \in \{d_1, d_2, d_3, d_4, d_5\}$. Then in d_1 and d_2, Z is the scoring rule winner. Indeed, for example, scores in d_1 are

$$\begin{aligned} s(X, d_1) &= 8a + 9b + 3c, \\ s(Y, d_1) &= 0a + 3b + 14c, \\ s(Z, d_1) &= 12a + 5b + 3c. \end{aligned}$$

Then, $s(Z, d_1) - s(X, d_1) = 4a - 4b > 0$ and $s(Z, d_1) - s(Y, d_1) = 12a + 2b - 14c > 0$.

On the other hand, in d_3, d_4 and d_5, Y is the scoring winner. In fact, for example, scores in d_3 are

$$\begin{aligned} s(X, d_3) &= 3a + 12b + 5c, \\ s(Y, d_3) &= 13a + 3b + 4c, \\ s(Z, d_3) &= 4a + 5b + 11c. \end{aligned}$$

Then $s(Y, d_3) - s(X, d_3) = 10a - 9b - c > 0$, $s(Y, d_3) - s(Z, d_3) = 9a - 2b - 7c > 0$.

preserved. That is, letting h be a non-decreasing function and $g_\ell = h(m - K_\ell)$, a proportional representation function with g also satisfies both consistency conditions.

Therefore, Z gets two seats, Y gets three seats, and X can not win in any district if any scoring rule is used. □

Proof of Theorem 1

Proof. Suppose that $S \geq 9$ and $E = (D, A, F)$ is a single-member district system. Without loss of generality we can assume

$$|d_1| \leq |d_2| \leq \cdots \leq |d_S|.$$

Let $k \equiv [\frac{1}{2}S] + 1$, where $[\cdot]$ is the floor function[*19].

Case 1. We first show (i) when F is Condorcet consistent or the plurality rule. Assume that $|d_k| \geq 3$ for each $k = 1, \ldots, S$. Let F be a Condorcet consistent rule or the plurality rule. We first show that there exists $\succsim \in \mathcal{R}_N$ such that there exists the Condorcet loser p_ℓ in P but p_ℓ is the dominant party, that is,

$$S_\ell(\succsim, E) \geq \left[\frac{S}{2}\right] + 1.$$

We shall construct a preference profile $\succsim \in \mathcal{R}^N$ as follows:

$\succsim_{d_j}$ $(j = 1, 2, \ldots, k)$		$\succsim_{d_j}$ $(j = k+1, k+2, \ldots, S)$
$\left[\frac{\lvert d_j\rvert}{2}\right] + 1$	$\lvert d_j\rvert - \left[\frac{\lvert d_j\rvert}{2}\right] - 1$	$\lvert d_j\rvert$
p_1	p_2	p_2
p_2	$\vdots$	$\vdots$
$\vdots$	p_1	p_1

Table 3: $\succsim$ for (i) when F is Condorcet consistent or the plurality rule.

For each district $j = 1, 2 \ldots, k$, since $\left[\frac{|d_j|}{2}\right] + 1 > \frac{|d_j|}{2}$, the Condorcet consistent rule or the plurality rule F chooses p_1, i.e., $F(\succsim_{d_j}) = (1, 0, 0, \ldots, 0)$. For each $j = k+1, k+2 \ldots, S$, the Condorcet consistent rule or the plurality rule F chooses p_2, i.e., $F(\succsim_{d_j}) = (0, 1, 0, \ldots, 0)$. Since p_1 wins $\left[\frac{S}{2}\right] + 1$ seats, p_1 is the dominant party. However, p_1 is the Condorcet loser in

[*19] $[x]$ is the largest integer not greater than x.

$\succsim$. To verify this, for each $p, q \in P$, define

$$\eta(p,q) \equiv |\{i \in N : p \succ_i q\}|.$$

It suffices to show that for all $q \in P$,

$$\eta(p_1, q) - \eta(q, p_1) < 0.$$

Take any $q \in P$. Then

$$\begin{aligned}
\eta(p_1,q) - \eta(q,p_1) &= \sum_{j=1}^{k}\left(\left[\frac{|d_j|}{2}\right]+1\right) - \left(\sum_{j=1}^{k}\left(|d_j| - \left[\frac{|d_j|}{2}\right] - 1\right) + \sum_{j=k+1}^{S}|d_j|\right) \\
&= \sum_{j=1}^{k}\left(2\left[\frac{|d_j|}{2}\right] - |d_j| + 2\right) - \sum_{j=k+1}^{S}|d_j| \\
&\leq \sum_{j=1}^{k} 2 - \sum_{j=k+1}^{S}|d_j|. \qquad (2.12)
\end{aligned}$$

The last weak inequality holds because $2\left[\frac{x}{2}\right] - x$ is either -1 or 0 for all integer x. Since $|d_j| \geq 3$ for all $d_j \in D$,

$$\sum_{j=1}^{k} 2 - \sum_{j=k+1}^{S}|d_j| \leq \sum_{j=1}^{k} 2 - \sum_{j=k+1}^{S} 3. \qquad (2.13)$$

Note that equality of (2.12) holds if and only if $|d_j|$ is even number for each $j = 1, 2, \ldots, k$. Moreover, note that equality of (2.13) holds if and only if $|d_j| = 3$ for all $j = k+1, \ldots, S$. Thus, if equality of (2.13) holds, since $3 \leq |d_1| \leq |d_2| \leq \cdots \leq |d_S|$, $|d_j| = 3$ for all $j = 1, 2, \ldots, S$, i.e., these are odd numbers. Therefore, strict inequality must hold in (2.12) or (2.13). Hence,

$$\begin{aligned}
\eta(p_1,q) - \eta(q,p_1) &< \sum_{j=1}^{k} 2 - \sum_{j=k+1}^{S} 3 \\
&= 2\left(\left[\frac{S}{2}\right]+1\right) - 3\left(S - \left[\frac{S}{2}\right] - 1\right)
\end{aligned}$$

$$
\begin{aligned}
&= 5\left(\left[\frac{S}{2}\right]+1\right) - 3S \\
&\leq 0 \qquad\qquad (2.14)
\end{aligned}
$$

The weak inequality in (2.14) holds because $S \geq 9$.[*20] Therefore, p_1 is the Condorcet loser and the dominant party. It is easy to see that p_2 is the Condorcet winner but its number of winning seats is less than that of p_1. Thus, any single-member district system with Condorcet consistent voting rule or the plurality rule is not Condorcet consistent.

Case 2. Next, we show (i) when F is the Borda rule. Assume that $|d_j| \geq 3$ for each $j = 1, \ldots, S$. Let F be the Borda rule. We show that there exists a preference profile such that there exists a party p_1 is the Condorcet loser but it is the dominant party in the congress. We shall construct a preference profile $\succsim \in \mathcal{R}_N$ as follows (Table 4).

$\succsim_{d_j}$ $(j = 1, 2, \ldots, k)$			$\succsim_{d_j}$ $(j = k+1, k+2, \ldots, S)$		
$\lvert d_j\rvert - \left[\frac{\lvert d_j\rvert}{2}\right] - 1$	$\lvert d_j\rvert - \left[\frac{\lvert d_j\rvert}{2}\right] - 1$	$2\left[\frac{\lvert d_j\rvert}{2}\right] - \lvert d_j\rvert + 2$	$\left[\frac{\lvert d_j\rvert}{2}\right]$	$\left[\frac{\lvert d_j\rvert}{2}\right]$	$\lvert d_j\rvert - 2\left[\frac{\lvert d_j\rvert}{2}\right]$
p_1	p_m	p_1	p_m	p_2	p_2
p_2	p_{m-1}	p_2	p_{m-1}	p_3	p_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p_{m-1}	p_2	p_{m-1}	p_2	p_m	p_m
p_m	p_1	p_m	p_1	p_1	p_1

Table 4: A preference profile

In districts $d_1, d_2, \ldots, d_k$, the Borda rule chooses candidates of the party p_1, so $S_1(\succsim) = \left[\frac{S}{2}\right] + 1$, that is, p_1 is the dominant party. In districts $d_k + 1, \ldots, d_S$, if $|d_j|$ is an even number, $p_2, \ldots, p_m$ get the same Borda scores, so the Borda rule arbitrarily chooses one of them, except for p_1. We can easily show that p_2 is the Condorcet winner but its number of winning seats is strictly less than that of p_1. Since $|d_j| \geq 3$, by the same argument in Case 1, p_1 is the Condorcet loser. Therefore, any single-member district electoral system with Borda rules is not Condorcet consistent.

Case 3. Next, we show (ii) when F is Condorcet consistent, the plurality rule or the Borda rule. Assume that $|d_j| \geq 8$ for each $j = 1, \ldots, S$. Let F be a Condorcet consistent rule or the

[*20]In fact, when $S = 5$ and 7, the equality holds. But when $S = 6$ and 8, it does not hold.

plurality rule or the Borda rule. We show that there exists a preference profile such that there exists a party p_1 is the Borda loser but it is the dominant party in the congress. We use again the preference profile $\succsim \in \mathcal{R}_N$ of Table 4.

In districts $d_1, d_2, \ldots, d_k$, candidates of the party p_1 win, so p_1 is the dominant party. We can see that the party p_m gets the least Borda score except for p_1. Therefore, to verify that p_1 is the Borda loser, it suffices to show that

$$B_1(\succsim) - B_m(\succsim) < 0.$$

By simple calculations, we have the following inequalities.

$$\begin{aligned}
B_1(\succsim) - B_m(\succsim) &= \sum_{j=1}^{k} B_1(\succsim_{d_j}) + \sum_{j=k+1}^{S} B_1(\succsim_{d_j}) - \sum_{j=1}^{k} B_m(\succsim_{d_j}) - \sum_{j=k+1}^{S} B_m(\succsim_{d_j}) \\
&= \sum_{j=1}^{k}(B_1(\succsim_{d_j}) - B_m(\succsim_{d_j})) + \sum_{j=k+1}^{S} (B_1(\succsim_{d_j}) - B_m(\succsim_{d_j})) \\
&= \sum_{j=1}^{k}(m-1)\left(2\left[\frac{|d_j|}{2}\right] - |d_j| + 2\right) + \sum_{j=k+1}^{S} |d_j| \\
&\quad - \sum_{j=k+1}^{S}\left(m\left[\frac{|d_j|}{2}\right] + 2\left(|d_j| - \left[\frac{|d_j|}{2}\right]\right)\right) \\
&= \sum_{j=1}^{k}(m-1)\left(2\left[\frac{|d_j|}{2}\right] - |d_j| + 2\right) - \sum_{j=k+1}^{S} m\left[\frac{|d_j|}{2}\right] \\
&\quad + \underbrace{\sum_{j=k+1}^{S}\left(2\left[\frac{|d_j|}{2}\right] - |d_j|\right)}_{\leq 0} \\
&\leq \sum_{j=1}^{k}(m-1)\left(2\left[\frac{|d_j|}{2}\right] - |d_j| + 2\right) - \sum_{j=k+1}^{S} m\left[\frac{|d_j|}{2}\right].
\end{aligned}$$

Since $2\left[\frac{|d_j|}{2}\right] - |d_j| + 2$ equals to 1 or 2,

$$\sum_{j=1}^{k}(m-1)\left(2\left[\frac{|d_j|}{2}\right] - |d_j| + 2\right) - \sum_{j=k+1}^{S} m\left[\frac{|d_j|}{2}\right] \leq \sum_{j=1}^{k} 2(m-1) - \sum_{j=k+1}^{S} m\left[\frac{|d_j|}{2}\right]$$

$$\leq \sum_{j=1}^{k} 2(m-1) - \sum_{j=k+1}^{S} m \left[\frac{|d_{k+1}|}{2}\right],$$

because $|d_{k+1}| \leq |d_{k+2}| \leq \cdots \leq |d_S|$.

Let $a = \left[\frac{|d_{k+1}|}{2}\right]$. Then, we have

$$\begin{aligned}
B_1(\succsim) - B_m(\succsim) &\leq \sum_{j=1}^{k} 2(m-1) - \sum_{j=k+1}^{S} ma \\
&= 2\left(\left[\frac{S}{2}\right] + 1\right)(m-1) - \left(S - \left[\frac{S}{2}\right] - 1\right) ma \\
&= \left(2\left[\frac{S}{2}\right] + 2 - Sa - \left[\frac{S}{2}\right] a - a\right) m \underbrace{-2\left(\left[\frac{S}{2}\right] + 1\right)}_{<0} \\
&< \left(2\left[\frac{S}{2}\right] + 2 - Sa - \left[\frac{S}{2}\right] a - a\right) m.
\end{aligned}$$

If

$$2\left[\frac{S}{2}\right] + 2 - Sa - \left[\frac{S}{2}\right] a - a \leq 0, \tag{2.15}$$

we can conclude $B_1 - B_m < 0$. When

$$|d_{k+1}| \geq \frac{4(S+3)}{S-3},$$

the inequality (2.15) holds. Since $S \geq 9$,

$$\frac{4(S+3)}{S-3} \leq \frac{48}{6} = 8.$$

Thus, by the assumption $|d_j| \geq 8$, (2.15) holds. Therefore, p_1 is the Borda loser in $\succsim$ but is the dominant party. We can easily show that p_2 is the Borda winner in $\succsim$. Hence, any single-memebr electoral system with a Condorcet consistent rule or the plurality rule or the Borda rule is not Borda consistent. □

Proof of Theorem 2

Proof. Let $E \in \mathcal{E}^p$ be a Borda proportional representation system, that is, $|D| = \{N\}$, $A(N) = S$, and F is a Borda proportional voting rule. Consider any $\succsim \in \mathcal{R}_N$. Note that

$$F_\ell(\succsim) = \left[\frac{B_\ell(\succsim)}{\lambda}\right]_r,$$

and

$$\sum_{p_\ell \in P} F_\ell(\succsim) = S$$

where $\lambda \in \mathbb{R}_{++}$ and $[\cdot]_r$ is some rounding function. By the following Lemma 1, the Borda rule ranks a Condorcet winner higher than a Condorcet loser, if p_ℓ is a Condorcet winner and p_j is a Condorcet loser,

$$B_\ell(\succsim) \geq B_j(\succsim).$$

Moreover, by the definition of the Borda rule, if p_ℓ is a Borda winner and p_j is a Borda loser,

$$B_\ell(\succsim) \geq B_j(\succsim).$$

Since $[\cdot]_r$ is nondecreasing,

$$F_\ell(\succsim) \geq F_j(\succsim).$$

Therefore, it is Condorcet and Borda consistent. □

Lemma 1. *The Borda rule ranks a Condorcet winner higher than a Condorcet loser.*

Proof. Consider any $\succsim \in \mathcal{R}$. Suppose that $p_w \in P$ is a Condorcet winner in $\succsim$ and $p_\ell \in P$ is a Condorcet loser in $\succsim$. We show that $B_w(\succsim) > B_\ell(\succsim)$.

Since p_w is a Condorcet winner,

$$|\{i \in N : p_w \succsim_i p_j\}| \geq \tfrac{1}{2}n \quad \text{for all } p_j \in P, \text{ and}$$
$$|\{i \in N : p_w \succsim_i p_k\}| > \tfrac{1}{2}n \quad \text{for som } p_k \in P.$$

Thus,

$$\sum_{p_j \in P} |\{i \in N : p_w \succsim_i p_j\}| > \frac{1}{2}nm.$$

Since $\sum_{p_j \in P} |\{i \in N : p_w \succsim_i p_j\}| = \sum_{i \in N} |\{p_j \in P : p_w \succsim_i p_j\}| = B_w(\succsim)$,

$$B_w(\succsim) > \frac{1}{2}nm.$$

On the other hand, by symmetry, we have $B_\ell(\succsim) < \frac{1}{2}nm$. Therefore, $B_w(\succsim) > B_\ell(\succsim)$. $\square$

Chapter 3

Multidimensional Evaluation: An Ordinal Approach

3.1 Introduction

Since Amartya Sen (1985) advocated capability approach, we have had a wide agreement that a multidimensional approach is necessary for measurement of human well-being. Researchers and policymakers have created many multidimensional measures such as the Human Development Index (HDI), Human Poverty Index (HPI), and Alkire-Foster Method (Alkire and Foster 2011a). Usually, achievements in different dimensions are not easily comparable because capabilities and functionings are ordinal in nature. Moreover, achievement in each dimension has its own value for capabilities (Sen 1990), so those should not be easily substitutable. Therefore, if we respect the idea of capabilities, we should respect incomparability and non-substitutability across different dimensions. However, currently used measures, particularly HDI and HPI, do not respect these ideas because these measures *carelessly* aggregate achievements in different dimensions.

The Human Development Index, created by the United Nations Development Programme in 1990, is a summary measure of three key dimensions of human development; income, health, and education. The current HDI aggregates achievements in three dimensions by the geometric mean. That is, it is of the form of

$$HDI = (y_1 \times y_2 \times y_3)^{\frac{1}{3}},$$

where y_i $(i = 1, 2, 3)$ denotes achievement in income, education, and health, respectively. This aggregation formula has some desirable properties and is characterized by a set of axioms (Herrero, Martínez, and Villar 2010; Zambrano 2014; Kawada, Nakamura, and Otani 2018). In this , , we claim that the current aggregation formula of HDI is not sophisticated because it cannot meet the idea of capabilities in the following sense. Since the aggregation formula is the geometric mean, achievements in different dimensions are treated easily comparable and substitutable: 1% increase in income is equally counted to 1% increase in human life. Needless to say, there is no rigor basis for comparing values in different dimensions and this way of comparisons does not respect the idea of capabilities.[*21]

Alkire and Foster (2011a) propose an approach for measurement of multidimensional poverty that can respect incomparability across dimensions, so-called the *Alkire-Foster method*. However, the Alkire-Foster method that respects cross-dimensional incomparability (M_0) cannot satisfy a traditional monotonicity condition: An index should increase if achievements in all dimensions increase.[*22] The traditional monotonicity condition has been emphasized in the literature since the seminal work of Sen (1976). Of course, HDI and many poverty measures except for the Alkire-Foster's M_0 satisfy the monotonicity condition. Summarizing above arguments, any existing multidimensional measure that respect cross-dimensional incomparability violates the traditional monotonicity.

In this chapter, we propose new methods for multidimensional evaluation that respects the monotonicity and cross-dimensional incomparability. We propose a new axiom *dimensional independence* that captures incomparability across dimensions. Then, we introduce new evaluation methods and show that our methods satisfy both of *monotonicity* and *dimensional independence*. Moreover, in a certain class of methods, we find a unique method that satisfies monotonicity, dimensional independence, and *minimal lower boundedness*. Our method can be applied to various multidimensional evaluation problems such that measurement of human development and multidimensional poverty. Using this method, we compute a new *human development indices* of 188 countries.

The rest of this chapter is organized as follows. In Section 3.2, we introduce our model. In Section 3.3, we introduce our methods and show main results. In Section 3.4, we apply our method for measurements of human development. In Section 3.5, we concludes this chapter.

[*21] In addition, the Human Poverty Index (HPI-1) is a summary measure of longevity, knowledge, and decent standard of livings. More precisely, it aggregates probability at birth of not surviving to age 40, adult illiteracy rate, ratio of population without sustainable access to an improved water source and children under weight for age (United Nations 2015). These different dimensions relate very different capabilities, but its aggregation formula do not sufficiently care about incomparability across dimensions.

[*22] More precisely, the Alkire-Fsoter Method is a class of measures M_α parametrized by $\alpha \geq 0$. In the class of the methods, M_0 works with ordinal data well and can respect cross-dimensional incomparability but it violates the monotonicity. On the other hand, when $\alpha > 0$, M_α satisfies the monotonicity but it cannot work with ordinal data well.

3.2 Definitions

Let $N = \{1, 2, \ldots, n\}$ be the set of nations and $D = \{1, 2, \ldots, d\}$ be the set of dimensions. We assume that $n \geq 3$ and $d \geq 3$. Let $\boldsymbol{y} = [y_{ik}] \in \mathbb{R}_+^{n \times d}$ denote the matrix of achievements, where y_{ik} is the achievement of nation i in dimension k. A row vector $y_i \in \mathbb{R}_+^d$ indicates the vector of achievements of nation i. Usually, some dimensions describe cardinal data and other dimensions do only ordinal data. Let $Y = \mathbb{R}_+^{n \times d}$ be the set of all possible data.

To make a ranking of nations, we introduce *score functions*. Let $s : N \times Y \to \mathbb{R}$ be a *score function*, where typical element $s(i, \boldsymbol{y})$ is the *score* of nation i when data is $\boldsymbol{y}$. A score should be interpreted as an ordinal number. Note that a score $s(i, \boldsymbol{y})$ may depend on data $\boldsymbol{y}$.

We introduce some axioms for score functions. We call a nation $i \in N$ *dominates* $j \in N$ at $\boldsymbol{y}$ if

$$
\begin{aligned}
y_{ik} &\geq y_{jk} \quad \text{for all} \quad k \in D, \ \textit{and} \\
y_{ik'} &> y_{jk'} \quad \text{for some} \quad k' \in D.
\end{aligned}
$$

The first axiom requires that a score of a nation should be higher than that of another nation if the former nation dominates the latter.

Axiom 1 (Monotonicity)**.** A score function s is *monotonic* if for each $\boldsymbol{y} \in Y$ and each $i, j \in N$,

$$s(i, \boldsymbol{y}) > s(j, \boldsymbol{y})$$

whenever i dominates j at $\boldsymbol{y}$.

Monotonicity seems to be weak, but some existing measures does not satisfy this. For example, the head-count ratio and Alkire-Foster methods M_0 violate monotonicity in general. Alkire-Foster methods M_α satisfy monotonicity when $\alpha > 0$ but they cannot work well with ordinal data. We introduce the following axiom that captures the idea “work well with ordinal data."

Axiom 2 (Dimensional independence)**.** A score function s is *dimensionally independent* if for each $\boldsymbol{y} \in Y$, each dimension $k \in D$, each strict increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$, and each

$i, j \in N$,

$$s(i, \boldsymbol{y}) \geq s(j, \boldsymbol{y}) \iff s(i, (f(\boldsymbol{y}_k), \boldsymbol{y}_{-k})) \geq s(j, (f(\boldsymbol{y}_k), \boldsymbol{y}_{-k})),$$

where $f(\boldsymbol{y}_k)$ is a column vector $(f(y_{1k}), f(y_{2k}), \ldots, f(y_{nk})) \in \mathbb{R}^N$.

The idea of this axiom has been often discussed in the literature (Atkinson 2003; Maasoumi and Lugo 2007; Alkire and Foster 2011a/b), but its formal definition has not been given before. If a score function is dimensionally independent, then it works with ordinal data well. The geometric mean, the current aggregation formula of the Human Development Index, violates dimensional independence. For example, $D = \{1, 2, 3\}$ and define $s(i, \boldsymbol{y}) = (y_{i1} y_{i2} y_{i3})^{\frac{1}{3}}$ for each $\boldsymbol{y} \in Y$ and $i \in N$. Let $\boldsymbol{y}_i = (4, 2, 1)$ and $\boldsymbol{y}_j = (8, 1, 1)$. Then $s(i, \boldsymbol{y}) = s(j, \boldsymbol{y}) = 8^{\frac{1}{3}} = 2$. However, let $f(x) = x^{\frac{1}{3}}$ and let $\boldsymbol{y}' = (f(\boldsymbol{y}_1), \boldsymbol{y}_{-1})$. Then

$$s(i, \boldsymbol{y}') = \left(4^{\frac{1}{3}} \times 2 \times 1\right)^{\frac{1}{3}} > 2^{\frac{1}{3}} = \left(8^{\frac{1}{3}} \times 1 \times 1\right)^{\frac{1}{3}} = s(j, \boldsymbol{y}').$$

Therefore, the geometric mean violates dimensional independence.

3.3 Methods and Results

We introduce a class of score functions that satisfy monotonicity and dimensional independence.

Definition 1 (score function $s*$)**.** Let s^* be the score function such that for each $\boldsymbol{y} \in Y$ and $i \in N$,

$$s^*(i, \boldsymbol{y}) = |\{j \in N : i \text{ dominates } j \text{ at } \boldsymbol{y}\}|.$$

The score function s^* of i simply counts the number of nations dominated by i. We can define a class of similar score functions generalizing the definition of the dominance relation. For each $t \in \{1, 2, \ldots, d\}$, each $\boldsymbol{y} \in Y$, and each $i, j \in N$, i *t-dominates* j at $\boldsymbol{y}$ if

$$|\{k \in D : y_{ik} \geq y_{jk}\}| \geq t \text{ and } |\{k \in D : y_{ik} > y_{jk}\}| \geq 1.$$

Definition 2 (score function s^t)**.** For each $t \in D$, let s^t be the score function such that for each $\boldsymbol{y} \in Y$ and $i \in N$,

$$s^t(i, \boldsymbol{y}) = |\{j \in N : i \text{ } t\text{-dominates } j \text{ at } \boldsymbol{y}\}|. \tag{3.16}$$

Note that $s^d = s*$. We call these score functions s^t $(t = 1, 2, \ldots, d)$ *simple score functions*. All simple score functions satisfy *monotonicity* and *dimensional independence*.

Proposition 1. *For each $t \in \{1, 2, \ldots, d\}$, the score function s^t satisfies* monotonicity *and* dimensional independence. *In particular, s^* satisfies both of them.*

Proof. See Appendix 3. □

In fact, there are many other score functions satisfy both monotonicity and dimensional independence. In particular, score functions s^B and s^N are intuitive examples of them.

Definition 3 (Borda score function s^B)**.** Let s^B be the score function such that for each $\boldsymbol{y} \in Y$ and $i \in N$,

$$s^B(i, \boldsymbol{y}) = \sum_{k \in D} |\{j \in N : y_{ik} \geq y_{jk}\}|.$$

In this definition, $|\{j \in N : y_{ik} \geq y_{jk}\}|$ is the number of nations that is ranked lower than i, namely, it is called the *Borda score* of i in dimension k (Borda 1784). Therefore, we call s^B the *Borda score function*. Similarly, we can define another score function multiplying Borda scores instead of summing them.

Definition 4 (Borda-Nash score function s^N)**.** Let s^N be the score function such that for each $\boldsymbol{y} \in Y$ and $i \in N$,

$$s^N(i, \boldsymbol{y}) = \prod_{k \in D} |\{j \in N : y_{ik} \geq y_{jk}\}|.$$

In this definition, it does not simply compute the product of $(y_{ik})_{k \in D}$, but does the product of the Borda scores of $\boldsymbol{y}_i$. That is the reason why we call s^N the *Borda-Nash score function*. In these definitions, s^B and s^N use only ordinal information of $\boldsymbol{y}$. Therefore, they satisfy dimensional independence.

Proposition 2. *The score functions s^B and s^N satisfy* monotonicity *and* dimensional independence.

Proof. See Appendix 3. □

We give a numerical example of ways of computation of these score functions. Consider that we have the following data $\boldsymbol{y}$ (Table 5). To compute scores easily, we make the matrix

	dimensions		
nations	1	2	3
1	28	5	3
2	30	4	0
3	25	7	2
4	25	5	1
5	32	6	3
6	25	3	1

Table 5: Numerical example: data $\boldsymbol{y}$

	dimensions			methods				
nations	1	2	3	s^*	s^1	s^2	s^B	s^N
1	4	4	6	2	4	4	14	96
2	5	2	1	0	4	1	8	10
3	3	6	4	2	5	3	13	72
4	3	4	3	1	2	2	10	36
5	6	5	6	4	5	5	17	180
6	3	1	3	0	0	0	7	9

Table 6: Numerical example: values of $|\{j \in N : y_{ik} \geq y_{jk}\}|$ and scores

whose component is $|\{j \in N : y_{ik} \geq y_{jk}\}|$ for each dimension $k \in D$ (Table 6).

For example, if we use s^*, nation 5 is ranked first and nation 2 and nation 6 are ranked worst.

There are many score functions satisfying monotonicity and dimensional independence. Then, which score function is the most appropriate for measuring the level of human development? Looking at the Table 5 in the above example, the achievement of nation 2 in dimension 3 is 0. However, scores of nation 2 is not the worst one in all score functions except for s^*. Thus, s^B, s^N, and s^t except for s^* violate *minimal lower boundedness*, a key axiom for measures of human development (Herrero, Martinez and Villar 2010).

Axiom 3 (Minimal Lower Boundedness)**.** A score function s is *minimal lower bounded* if for each $\boldsymbol{y} \in Y$ and each $i, j \in N$,

$$s(i, \boldsymbol{y}) \leq s(j, \boldsymbol{y})$$

whenever $\min(\boldsymbol{y}_i) = 0 < \min(\boldsymbol{y}_\ell)$ for any $\ell \in N \setminus \{i\}$.

Proposition 3. *The score functions s^B and s^N violate minimal lower boundedness.*

Proof. In the above example (Tables 5 and 6), the achievement of nation 2 in dimension 3 is 0, and any other achievement in the matrix is larger than 0. However, $s^B(2, \boldsymbol{y}) = 8 > 7 = s^B(6, \boldsymbol{y})$ and $s^N(2, \boldsymbol{y}) = 10 > 9 = s^N(6, \boldsymbol{y})$. Therefore, s^B and s^N violate minimal lower boundedness. □

Finally, we state that s^* is the only method that satisfy minimal lower boundedness in the class of simple score functions.

Proposition 4. *For any $t = 1, 2, \ldots, d$, if a score function s^t satisfies minimal lower boundedness, then $s^t = s^*$.*

Proof. See Appendix 3. □

3.4 New HDI Ranking

Using s^*, we make a new HDI ranking (Figure 3).[*23] We compute the ranking of 188 countries but we only display the ranking of the top 18 countries here. The ranking $R1$ is the current HDI ranking and $R2$ is the new ranking based on s^*. In fact, these rankings are very similar. For example, the cosine similarity between two rankings (rankings of 188 countries) is 0.9965. [*24]

However, there exist some countries such that their rankings in $R1$ are very different from those in $R2$. In particular, Kuwait's and Singapore's rankings saliently differ in current and new HDI rankings (Figure 4). The ranking of Kuwait in $R1$ is 51st but that in $R2$ is 90th, so $R1 - R2$ is -39. To see the reason why R1 and R2 of Kuwait are so different, we look at the data of Kuwait and similarly-ranked countries (Figure 5). In Figure 5, GNI per capita of Kuwait is relatively very high among those of similarly-ranked countries. Then, low achievements in other dimensions of Kuwait are substituted by the high achievement in the income dimension. Therefore, the ranking of Kuwait is computed as too high in $R1$.

[*23]Data are sourced from the United Nations Development Report 2015 (United Nations 2015).

[*24]Cosine similarity is a commonly used similarity measure in computer science discussed in Chapter 1. The definition of the cosine similarity between two vectors is as follows: For each vector $x, y \in \mathbb{R}^m_{++}$, *cosine similarity* between x and y is

$$C(x, y) \equiv \frac{x \cdot y}{\|x\|\|y\|},$$

where $\|x\|$ is the Euclidean norm of x, and $x \cdot y$ denotes the inner product between x and y.

That is, the current aggregation method of HDI (the geometric mean) might treat values in different dimensions substitutable. We claim that this way of computation of HDI does not meet the idea of capabilities. That is mainly because the current HDI does not satisfy *dimensional independence*. On the other hand, in $R2$, based on our method s^*, comparability and substitutability across dimensions are almost excluded because s^* satisfies *dimensional independence*. Therefore, if we respect capability approach, then we should use s^* for computing HDI ranking.

HDI Rank (R1)	Countrty	Our New HDI Rank (R2)	R1 - R2
1	Norway	2	-1
2	Australia	1	1
2	Switzerland	8	-6
4	Germany	4	0
5	Denmark	7	-2
5	Singapore	19	-14
7	Netherlands	12	-5
8	Ireland	3	5
9	Iceland	6	3
10	Canada	9	1
10	United States	14	-4
12	Hong Kong	21	-9
13	New Zealand	4	9
14	Sweden	9	5
15	Liechtenstein	27	-12
16	United Kingdom	12	4
17	Japan	16	1
18	Korea (Republic of)	9	9

Figure 3: Current ranking and new ranking of HDI

	R1 (current HDI)	R2 (new HDI)	R1 - R2
Kuwait	51st	90th	-39
Singapore	5th	19th	-14

Figure 4: Kuwait's and Singapore's rankings saliently differ in $R1$ and $R2$.

R1		HDI	Life expectancy	Expected years of schooling	Mean years of schooling	GNI per capita
48	Montenegro	0.807	76.4	15.1	11.3	15410
49	Russian Federatior	0.804	70.3	15.0	12.0	23286
50	Romania	0.802	74.8	14.7	10.8	19428
51	Kuwait	0.800	74.5	13.3	7.3	76075
52	Belarus	0.796	71.5	15.7	12.0	15629
52	Oman	0.796	77.0	13.7	8.1	34402
54	Barbados	0.795	75.8	15.3	10.5	14952

Figure 5: Data of Kuwait and similarly-ranked countries

3.5 Conclusion

We have proposed new methods for multidimensional evaluation when achievements in different dimensions are not easily comparable or sustitutable. Existing methods for multidimensional evaluation do not simultaneously satisfy *monotonicity* and *dimensional independence*, but our proposed methods satisfy both of them. Moreover, in a certain class of methods, we have found a unique method that satisfies *monotonicity*, *dimensional independence*, and *minimal lower boundedness*. Our methods can be applied to various problems of multidimensional evaluation such as measurements of human development and multidimensional poverty. In this chapter, we have applied one of our methods to make new HDI ranking of 188 countries and compare the current and new HDI rankings. According to the idea of capability approach, which is the basic concept of HDI, different achievements in different dimensions should not be easily substitutable. Comparing the current and new HDI rankings, we have claimed that the current HDI does not respect the property. Finally, we have claimed that if we respect the idea of capabilities, we should use our measure for making HDI rankings. Applying our methods to measurement of multidimensional poverty remains in future research.

Appendix 3

Proof of Proposition 1

Lemma 1. *For any $\boldsymbol{y} \in Y$, each $i, j \in N$, and each $t \in \{1, 2, \ldots, d\}$, if i dominates j at $\boldsymbol{y}$, then i t-dominates j at $\boldsymbol{y}$.*

Proof. Take any $\boldsymbol{y} \in Y$, $i, j \in N$, and $t \in \{1, 2, \ldots, d\}$. Suppose that i dominates j at $\boldsymbol{y}$. Then

$$y_{ik} \geq y_{jk} \text{ for all } k \in D \text{ and } y_{ik'} > y_{jk'} \text{ for some } k' \in D.$$

Therefore,

$$|\{k \in D : y_{ik} \geq y_{jk}\}| = d \geq t \text{ and } |\{k' \in D : y_{ik'} > y_{jk'}\}| \geq 1.$$

Thus, i t-dominates j at $\boldsymbol{y}$. □

Lemma 2. *The dominance relation is transitive. That is, for any $\boldsymbol{y} \in Y$, and each $i, j, \ell \in N$, if i dominates j at $\boldsymbol{y}$ and j dominates ℓ at $\boldsymbol{y}$, then i dominates ℓ at $\boldsymbol{y}$.*

Proof. Consider any $\boldsymbol{y} \in Y$, and any $i, j, \ell \in N$ such that i dominates j at $\boldsymbol{y}$ and j dominates ℓ at $\boldsymbol{y}$. Then

$$y_{ik} \geq y_{jk} \geq y_{\ell k}$$

for all $k \in D$. Moreover, $y_{ik'} > y_{jk'}$ or $y_{jk'} > y_{\ell k'}$ holds for some $k' \in D$. Thus, i dominates ℓ at $\boldsymbol{y}$. $\square$

Proof of Proposition 1. Consider any $t \in \{1, 2, \ldots, \}$. Let s^t be the score function defined by (3.16).

We first show that s^t satisfies monotonicity. Consider any $\boldsymbol{y} \in Y$ and take any $i, j \in N$ such that i dominates j at $\boldsymbol{y}$. By Lemmas 1 and 2, it has been shown that for any $\ell \in N$ such that ℓ is t-dominated by i at $\boldsymbol{y}$, ℓ is also t-dominated by j at $\boldsymbol{y}$. Moreover, since j itself is t-dominated by i at $\boldsymbol{y}$,

$$|\{\ell \in N : i \text{ } t\text{-dominates } \ell \text{ at } \boldsymbol{y}\}| \geq |\{\ell \in N : j \text{ } t\text{-dominates } \ell \text{ at } \boldsymbol{y}\}| + 1.$$

Therefore,

$$\begin{aligned} s^t(i, \boldsymbol{y}) &= |\{\ell \in N : i \text{ } t\text{-dominates } \ell \text{ at } \boldsymbol{y}\}| \\ &> |\{\ell \in N : j \text{ } t\text{-dominates } \ell \text{ at } \boldsymbol{y}\}| \\ &= s^t(j, \boldsymbol{y}). \end{aligned}$$

That is, s^t satisfies monotonicity.

Next, we show that s^t satisfies dimensional independence. Let f be any strict increasing function $\mathbb{R}_+ \to \mathbb{R}_+$. Consider any $\boldsymbol{y} \in Y$ and any $k \in D$. Let $\boldsymbol{y}' = (f(\boldsymbol{y}_k), \boldsymbol{y}_{-k})$. It suffices to show that $s^t(i, \boldsymbol{y}) = s^t(i, \boldsymbol{y}')$ for any $i \in N$. Since f is strict increasing function,

$$y_{ik} \geq y_{jk} \iff f(y_{ik}) \geq f(y_{jk}) \iff y'_{ik} \geq y'_{jk}$$

for all $i, j \in N$. Therefore, $s^t(i, \boldsymbol{y}) = s^t(i, \boldsymbol{y}')$ for any $i \in N$. $\square$

Proof of Proposition 2

Proof. Let s^B be the Borda score function. Consider any $\boldsymbol{y} \in Y$ and any $i, j \in N$. Suppose that i dominates j at $\boldsymbol{y}$. We show that $s^B(i, \boldsymbol{y}) > s^B(j, \boldsymbol{y})$.

By the definition of the dominance relation, for each $k \in D$ and each $\ell \in N$, if $y_{jk} \geq y_{\ell k}$, then $y_{ik} \geq y_{\ell k}$. Therefore, for each $k \in D$,

$$|\{\ell \in N : y_{ik} \geq y_{\ell k}\}| \geq |\{\ell \in N : y_{jk} \geq y_{\ell k}\}|.$$

Moreover, because j is dominated by i, there exists $k' \in D$ such that

$$|\{\ell \in N : y_{ik'} \geq y_{\ell k'}\}| \geq |\{\ell \in N : y_{jk'} \geq y_{\ell k'}\}| + 1.$$

Thus,

$$\sum_{k \in D} |\{\ell \in N : y_{ik} \geq y_{\ell k}\}| > \sum_{k \in D} |\{\ell \in N : y_{jk} \geq y_{\ell k}\}|,$$

that is, $s^B(i, \boldsymbol{y}) > s^B(j, \boldsymbol{y})$. It is easy to show that s^B is dimensionally independent because any strict increasing function f does not change values of $|\{\ell \in N : y_{ik} \geq y_{\ell k}\}|$ for all $i, \ell \in N$ and $k \in D$.

A proof for the Borda-Nash score function s^N is similar. $\square$

Proof of Proposition 4

Proof. We first show that s^* satisfies minimal lower boundedness. Consider any $\boldsymbol{y} \in Y$. Suppose that $\min(\boldsymbol{y}_i) = 0$ and $\min(\boldsymbol{y}_j) > 0$ for all $j \in N \setminus \{i\}$. Then, i cannot dominate any other nation. Therefore, $s^*(i, \boldsymbol{y}) = 0$. This implies that i is ranked worst.

Next, we show that s^t violates minimal lower boundedness if $t \neq d$. Take any $t \in \{1, 2, \ldots, d-1\}$ and let s^t be the simple score function. Let $\boldsymbol{y}_1 = (0, 1, 1, \ldots, 1) \in \mathbb{R}^d_+$ and $\boldsymbol{y}_j = (1, 0.1, 0.1, \ldots, 0.1) \in \mathbb{R}^d_+$ for each $j = 2, 3, \ldots, n$. Then, nation 1 t-dominates j at $\boldsymbol{y}$ for any $j = 2, 3, \ldots, n$. Therefore, $s^t(i, \boldsymbol{y}) > 1$ because $n \geq 3$ and $d \geq 3$. On the other hand, we first consider the case in which $t \neq 1$. Then, nation j cannot t-dominates ℓ for any $\ell \in N$. Therefore, $s(j, \boldsymbol{y}) = 0$ for each $j = 2, 3, \ldots, n$. We next consider the case in which $t = 1$. Then nation j only can 1-dominates 1. Therefore, $s(j, \boldsymbol{y}) = 1$ for each $j = 2, 3, \ldots, n$. Thus, the ranking of nation 1 is not worst one though it has the minimal achievement. Hence, s^t violates minimal lower boundedness. $\square$

Chapter 4

The Measurement of Population Ageing

[*25]

4.1 Introduction

Population ageing is one of the most serious problems in many developed countries. According to the United Nations Population Division Reports, it is unprecedented, pervasive, and enduring (United Nations 2002). The level of population ageing is often measured by the ratio of the older population among the entire population (e.g., people aged over 65 years), the so-called *head-count ratio*. For example, in Japan, the head-count ratio is 4.8% in 1950, 26.6% in 2015, and is projected to reach 33.4% by 2035.[*26] Although the head-count ratio is quite often used, it violates at least two elementary properties for measuring population ageing.

First, the head-count ratio violates a monotonicity property with respect to ages. For example, even though all older individuals become further older, as long as the number of them remains the same, the increase in ages of older individuals does not affect the head-count ratio. This means that even if the head-count ratio is 10%, a distribution with “10% of individuals mostly in the range 65–74 years old" cannot be distinguished with a distribution with“10% of individuals mostly in the range 75–84 years old".

Moreover, the index is insensitive to the distribution of ages. Especially, it fails to take thickness of the working age population into account (e.g., people aged between 20 and 64 years). One non-negligible facet of population ageing is that the supply of the labor force by the working age population becomes scarce to the demand by the older population or economically inactive population. Particularly, the government of Japan confronts with a substantial budget imbalance due to large increases in the public expenditures for social

[*25] This chapter is co-authored with Yuta Nakamura and Noriaki Okamoto, and based on Kawada, Nakamura, and Okamoto (2017).

[*26] The population structure data and projections in Japan are sourced from the National Institute of Population and Social Security Research 2017.

security systems and a decrease in tax-revenue caused by shrinking labor force, e.g., total expenditures for social secury systems exceed 23% of GDP in 2014 in Japan (Kitao 2015). However, since the head-count ratio simply counts the number of older people and checks the percentage among the entire population including the people aged under 19 years, it fails to reflect thickness of the working age population. Similarly, all measures used in the United Nations reports such as the median age and the total dependency ratio have at least one of these two drawbacks.

The choice of measures of population ageing is important. That is because it is difficult to capture the complicated phenomenon of population ageing without help of some measures, and they shape our perception of demographic trends. Therefore, if a government uses a “bad" measure to perceive demographic trends, the government might misperceive the trends and fail to plan appropriate policies against population ageing.

In this chapter, we propose a new measure of population ageing that overcomes shortcomings of the measures currently in use. We characterize the new measure by *monotonicity*, the *working age principle*, and other standard axioms. The *working age principle*, which is introduced in this ， and inspired by the *Pigou-Dalton transfer principle* (Dalton 1920), is a sensitivity condition to thickness of the working age population. We also compute our measure for population data of China and Japan. This computation illustrates differences between our measure and the head-count ratio.

Our measure satisfies both of two elementary properties, monotonicity and the working age principle, and existing measures violate at least one of these properties. In this sense, our measure improves existing measures. Of course, it does not mean that our measure supersedes all of them. Population ageing is a complex phenomenon with various facets, so we cannot fully measure it using just one measure. Our contribution is to add a new measure to a set of tools for measuring population ageing.

The rest of this study is organized as follows. Section 4.2 introduces our model. Section 4.3 presents our axioms. In Section 4.4, we characterize the new index function of population ageing by a set of axioms. Section 4.5 concludes this chapter. All omitted proofs are relegated to Appendix.

4.2 Model

Let $N = \{1, \ldots, n\}$ be a set of individuals in a society and $n \geq 3$. Each individual $i \in N$ has a corresponding age $y_i \in [0, \bar{y}] \subset \mathbb{R}$. An *age profile* is a list of individual ages $\boldsymbol{y} = (y_1, \ldots, y_n) \in [0, \bar{y}]^n$. An age $y_i \in [0, \bar{y}]$ is a *working age* if $y_i \in [x, z]$, where $x, z \in (0, \bar{y}]$ and $x < z$. The age x indicates the requirement age for starting work, and z does the retirement age. The *residual term for working* at $y_i \in [0, \bar{y}]$ is

$$a(y_i) = \begin{cases} z - x & \text{if } y_i < x, \\ z - y_i & \text{if } x \leq y_i \leq z, \\ 0 & \text{if } z < y_i, \end{cases}$$

that is, at age $y_i < x$, individual i has the residual term $z - x$ for working since he has not attained working ages, at age $x \leq y_i \leq z$, individual i has the residual term $z - y_i$ for working since he has worked for $y_i - x$ years, and at age $y_i > z$, individual i has no residual term for working since he has already passed the retirement age. An *index function* is a function $I : [0, \bar{y}]^n \to \mathbb{R}$ that maps each age profile $\boldsymbol{y} \in [0, \bar{y}]^n$ to a real number $I(\boldsymbol{y}) \in \mathbb{R}$. For example, an index function defined as

$$I^H(\boldsymbol{y}) = \frac{|\{i \in N : y_i > z\}|}{n}$$

is called the *head-count ratio*.

4.3 Axioms

Continuity requires that an index function be robust to small misspecification of data.

Continuity. An index function $I : [0, \bar{y}]^n \to \mathbb{R}$ is continuous.

We through this chapter assume that index functions are continuous.[*27] Next, *monotonicity* requires that if all individuals' ages weekly increase and some individuals' ages strictly increase, then the index strictly increase.

[*27]In our analysis, an age is a real number in $[0, \bar{y}]$ because we actually consider the normalized age $y_i/\bar{y} \in [0, 1]$ in characterization results.

Monotonicity. For each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$, if $\boldsymbol{y}_i \geq \boldsymbol{y}'_i$ for all $i \in N$ and $\boldsymbol{y}_j > \boldsymbol{y}'_j$ for some $j \in N$, then $I(\boldsymbol{y}) > I(\boldsymbol{y}')$.

Consider two arbitrary age profiles with a common age for some individual. *Separability* requires that if this common age is replaced with another one, then an index function preserves the order between the two age profiles.

Separability. For each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$, and each $j \in N$,

$$I(y_j, \boldsymbol{y}_{-j}) \geq I(y_j, \boldsymbol{y}'_{-j}) \Longrightarrow I(y'_j, \boldsymbol{y}_{-j}) \geq I(y'_j, \boldsymbol{y}'_{-j}).$$

The next axiom is proposed in this , , which requires sensitivity to thickness of the working age population. The *working age principle* requires that for any age profile, if any two individuals' ages are replaced by others while preserving its sum, then an index function weakly decreases whenever the sum of residual terms for working among these two individuals weakly increases.

The Working Age Principle. For each $\boldsymbol{y} \in [0, \bar{y}]^n$, each $i, j \in N$, and each $y'_i, y'_j \in [0, \bar{y}]$ with $y'_i + y'_j = y_i + y_j$,

$$a(y_i) + a(y_j) \leq a(y'_i) + a(y'_j) \implies I(\boldsymbol{y}) \geq I(y'_i, y'_j, \boldsymbol{y}_{-\{i,j\}}).$$

This axiom is inspired by the *Pigou-Dalton transfer principle* (Dalton 1920). The *working age principle* can be interpreted as follows: A virtual transfer of age from an individual to another individual weakly reduces the value of the index if the sum of their residual terms for working weakly increases. For example, consider the following three age profiles in which $n = 3$,

$$\begin{aligned} \boldsymbol{y} &= (25, 45, 75), \\ \boldsymbol{y}' &= (25, 50, 70), \\ \boldsymbol{y}'' &= (25, 40, 80). \end{aligned}$$

Suppose that $x = 20$ and $z = 65$ in this society. Since

$$a(y_2) + a(y_3) = (65 - 45) + 0 = 20 > 15 = (65 - 50) + 0 = a(y_2') + a(y_3'),$$

that is, a transfer of 5 age from individual 3 to individual 2 causes a decrease in the sum of their residual terms for working, it follows that $I(\boldsymbol{y}) \leq I(\boldsymbol{y}')$. On the other hand, since $a(y_2) + a(y_3) < a(y_2'') + a(y_3'')$, that is, a transfer of 5 age from individual 2 to 3 causes an increase in the sum of their residual terms for working, it follows that $I(\boldsymbol{y}) \geq I(\boldsymbol{y}'')$.

Finally, *normalization* requires that for any age profile, if all individuals are the same age, then its index takes a value of the age over the maximal age $\bar{y}$.

Normalization. For each $y \in [0, \bar{y}]$,

$$I(y, \ldots, y) = \frac{y}{\bar{y}} \in [0, 1].$$

4.4 A New Measure of Population Ageing

Our purpose is to search for measures of population ageing that satisfy elementary properties stated in the previous section. First, we introduce an index function that represents an ordering on age profiles. For each $\alpha \geq 0$, let $I_\alpha : [0, \bar{y}]^n \to \mathbb{R}$ be such that

$$I_\alpha(\boldsymbol{y}) = \sum_{j=1}^{n} y_j + \alpha \sum_{j=1}^{n} (z - x - a(y_j)).$$

We show that a continuous index function I satisfies *monotonicity*, *separability*, and the *working age principle* if and only if the ordering represented by I is the same ordering represented by I_α for some $\alpha \geq 0$.

Theorem 1. *For each continuous index function $I : [0, \bar{y}]^n \to \mathbb{R}$, the following statements (i) and (ii) are equivalent:*

(i) $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies monotonicity, separability, and the working age principle;

(ii) there exists $\alpha \geq 0$ such that for each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$,

$$I(\boldsymbol{y}') \geq I(\boldsymbol{y}) \iff I_\alpha(\boldsymbol{y}') \geq I_\alpha(\boldsymbol{y}).$$

Since I_α is a representation function of an ordering, the represented ordering remains unchanged if I_α is monotonically transformed. For example, let

$$I'_\alpha(\boldsymbol{y}) = \frac{1}{1+\alpha}\left(\frac{1}{n}\sum_{j=1}^{n} y_j + \alpha\frac{1}{n}\sum_{j=1}^{n}\Big(z - x - a(y_j)\Big)\right).$$

Then I'_α represents the same ordering. The first term in the biggest bracket in I'_α is the mean age, the second one is the mean term that individuals have worked for, and I'_α is a convex combination of them. A parameter α is the degree of sensitivity to the thickness of working age population. It could coincide with the mean age when $\alpha = 0$. This is caused by the weakness of the *working age principle*. Indeed, it permits that whenever $y'_i + y'_j = y_i + y_j$ holds,

$$I(\boldsymbol{y}) = I(y'_i, y'_j, \boldsymbol{y}_{-\{i,j\}}).$$

We introduce the *strict working age principle* to exclude the mean age, which fails to respect the thickness of working age population.

The Strict Working Age Principle. For each $\boldsymbol{y} \in [0, \bar{y}]^n$, each $i, j \in N$, and each $y'_i, y'_j \in [0, \bar{y}]$ with $y'_i + y'_j = y_i + y_j$,

$$\begin{aligned} a(y_i) + a(y_j) = a(y'_i) + a(y'_j) &\implies I(\boldsymbol{y}) = I(y'_i, y'_j, \boldsymbol{y}_{-\{i,j\}}), \\ a(y_i) + a(y_j) < a(y'_i) + a(y'_j) &\implies I(\boldsymbol{y}) > I(y'_i, y'_j, \boldsymbol{y}_{-\{i,j\}}). \end{aligned}$$

Replace the *working age principle* with the strict version, we have the following corollary.

Corollary 1. *For each continuous index function $I : [0, \bar{y}]^n \to \mathbb{R}$, the following statements (i) and (ii) are equivalent:*

(i) $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies monotonicity, separability, and the strict working age principle;

(ii) there exists $\alpha > 0$ such that for each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$,

$$I(\boldsymbol{y}') \geq I(\boldsymbol{y}) \iff I_\alpha(\boldsymbol{y}') \geq I_\alpha(\boldsymbol{y}).$$

Examples showing the tightness of the axioms in Theorem 1 and Corollary 1 are provided in the Appendix. Finally, we characterize an index function that satisfies *monotonicity*, *separability*, the *strict working age principle*, and the *normalization*.

For each $\alpha > 0$, let $f_\alpha : [0, \bar{y}] \to \mathbb{R}$ be such that for each $y_i \in [0, \bar{y}]$,

$$f_\alpha(y_i) = \begin{cases} y_i & \text{if } y_i < x, \\ (1+\alpha)y_i - \alpha x & \text{if } x \leq y_i \leq z, \\ y_i + \alpha(z - x) & \text{if } z < y_i. \end{cases}$$

Note that

$$I_\alpha(\boldsymbol{y}) = \sum_{j=1}^{n} f_\alpha(y_j).$$

Corollary 2. *For each continuous index function* $I : [0, \bar{y}]^n \to \mathbb{R}$*, the following statements (i) and (ii) are equivalent:*

(i) $I : [0, \bar{y}]^n \to \mathbb{R}$ *satisfies monotonicity, separability, the strict working age principle, and* normalization*;*

(ii) there exists $\alpha > 0$ *such that*

$$I(\boldsymbol{y}) = \frac{1}{\bar{y}} \cdot f_\alpha^{-1}\left(\frac{1}{n} I_\alpha(\boldsymbol{y})\right).$$

We illustrate differences between the head-count ratio and our new measure, using China's and Japan's population data.*28 We focus on China's 1990–2015 data for three reasons: (i) China's population ageing had became serious in this term. Indeed, in 2002, China was classified as an "ageing society" by the United Nations since its head-count ratio exceeded 7%. (ii) Moreover, the year 1990 is about ten years after the government of China officially enacted its unparalleled "one-child policy". (iii) In addition, at almost the same time, market-oriented economic reforms were made by the government, which induced several decades of rapid economic growth that would also tend to decrease fertility rates in China (Zhang 2017). By focusing on this term, we check policy impacts on population ageing in China.

We use Japan's data as a benchmark since its population ageing is the most serious one. We focus on Japan's 1962–1987 data because in this term, also in Japan, its population ageing

*28 Population data are sourced from United Nations (2017).

had became serious and an economic growth called the “Japanese economic miracle" had occurred. In addition, we focus on these terms since the head-count ratio and our measure behave very differently in the following computation.

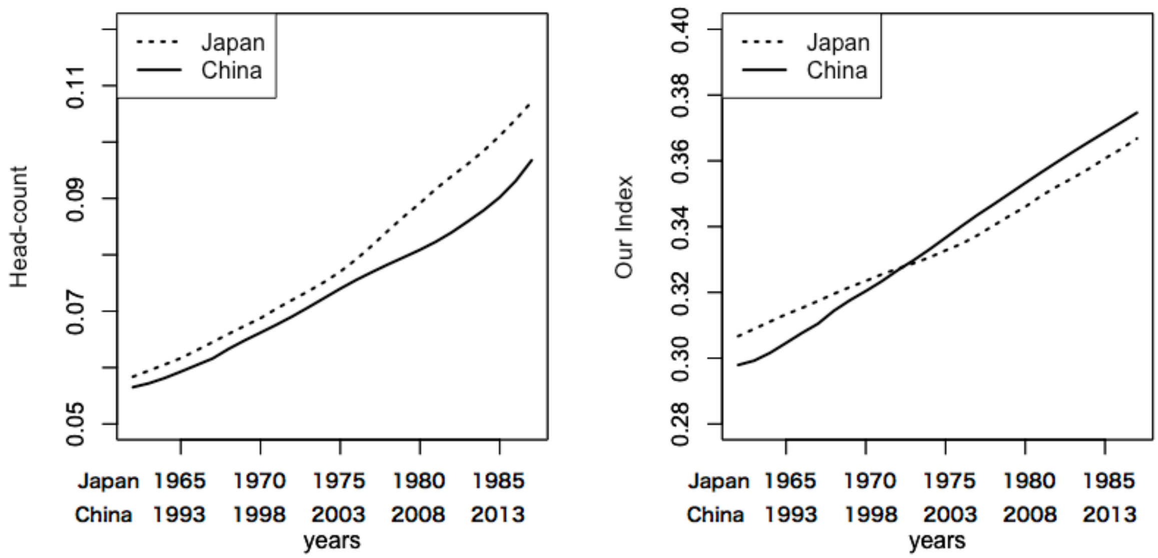

Figure 6: the head-count ratios and our measures

Figure 6 shows computation results of the head-count ratios and our measures for China’s 1990–2015 and Japan’s 1962–1987 data. In the left plot, we can see that the head-count ratio of China increases in a similar way as Japan in these terms, but it is always smaller than that of Japan. So, we use the head-count ratio as a measure of population ageing, we perceive that China’s population ageing is similar to Japan in these terms. However, it may be misleading. In the right plot, we can see that our measure for China, rapidly increases and exceeds that for Japan in the same terms.[*29]

This empirical example tells us the following things. If we use the head-count ratio, we might underestimate the population ageing in China. On the other hand, our measure can

[*29] In this computation, we fix $x = 20$, $z = 65$, $\bar{y} = 100$, and $\alpha = 1$ for simplicity. However, we can get similar results if we change these exogenous variables to some extent. For example, when we change $x_{China} = 15$ while x_{Japan} remains the same by considering conventions for starting age for working in China, our measure for China will decrease because the amount of labor force will increase, but we can get a similar result in which our measure for China exceeds that for Japan.

vividly capture China's rapid population ageing. This is because our measure is sensitive to the thickness of potential working age population more than the head-count ratio. Our measure for China sharply increases probably because of a decrease in fertility rate affected by governmental policies of China.

4.5 Concluding Remarks

As far as the authors know, this study is the first one that axiomatically analyzes the measurement of population ageing. Our study is inspired by Sen (1976)'s criticism of the head-count ratio in measurement of poverty. A distinct point from the literature of measurement of poverty is that they focus on left-tail of income distributions (low incomes) and require monotonicity and sensitivity to inequality within left-tail distributions (e.g., Sen 1976; Foster and Shorrocks 1991). Differing from poverty, population ageing is a trend on the entire distribution, so using only information of right-tail distributions and discarding residual information is inadequate to measure the level of population ageing. Therefore, we do not focus on right-tail of age distributions (older populations). That point mainly differs from Chu (1997), which is inspired by the literature of measurement of poverty and proposes a new measure of population ageing, too. Indeed, his measure focuses on the right-tail of age distributions. In addition, he does not provide an axiomatization of his measure.

In our analysis, we treat the size of population as fixed. This assumption is imposed only for simplifying notation and statements of the axioms. Indeed, we can straightforwardly extend all of our results by invoking results by Foster and Shorrocks (1991). Foster and Shorrocks (1991) consider income distributions with variable population sizes to characterize *subgroup consistent* poverty indices. Subgroup consistency can be interpreted as a generalized version of our separability, and they characterize the orderings of *canonical indices* by using it. Therefore, we can apply their results to characterize a modified version of our index functions I_α. In the characterization, the modified index function would be characterized by a modified version of the working age principle and their set of axioms.

Our proposed measure is the unique measure that satisfies monotonicity, separability, the strict working age principle, and normalization. Our measure improves all existing measures in this sense. However, we do not claim that our measure supersedes all of them. Population ageing is a complex phenomenon with various facets, so we need to use various measures to capture it. Our contribution is to add a new tool to a set of measures for population ageing.

Appendix 4

In our proof of Theorem 1, we apply Debreu's (1959) representation theorem of a preference on a separable set of variables. We introduce some definitions to apply Debreu's theorem.

Debreu considers a continuous and complete preordering $\succsim$ on a commodity-bundle space $S \subset \mathbb{R}^{\ell}$. Suppose that this space can be decomposed into n subspaces $S_1, \ldots, S_n$ ($n \leq \ell$), that is,

$$S = \times_{i=1}^{n} S_i.$$

The factors $1, \ldots, n$ are *independent* if for all $i \in \{1, \ldots, n\}$, all $\boldsymbol{y}_i, \boldsymbol{y}'_i \in S_i$, and all $\boldsymbol{y}_{-i}, \boldsymbol{y}'_{-i} \in \times_{j \neq i} S_j$,

$$(\boldsymbol{y}_i, \boldsymbol{y}_{-i}) \succsim (\boldsymbol{y}'_i, \boldsymbol{y}_{-i}) \iff (\boldsymbol{y}_i, \boldsymbol{y}'_{-i}) \succsim (\boldsymbol{y}'_i, \boldsymbol{y}'_{-i}).$$

A factor $i \in \{1, \ldots, n\}$ is *essential* if there exist $\boldsymbol{y}_{-i} \in \times_{j \neq i} S_j$ and $\boldsymbol{y}_i, \boldsymbol{y}'_i \in S_i$ such that

$$(\boldsymbol{y}_i, \boldsymbol{y}_{-i}) \succ (\boldsymbol{y}'_i, \boldsymbol{y}_{-i}).$$

Debreu's Representation Theorem (Debreu 1959, Theorem 3)**.** Suppose that the factors $1, \ldots, n$ are independent and that at least three of them are essential. If S_i is connected for each $i \in \{1, \ldots, n\}$, then a continuous and complete preordering $\succsim$ on $S = \times_{i=1}^{n} S_i$ can be represented by an additively separable utility function: that is, for each $i \in \{1, \ldots, n\}$, there exists a continuous function $U_i : S_i \to \mathbb{R}$, and for each $\boldsymbol{y}, \boldsymbol{y}' \in S$,

$$\boldsymbol{y} \succsim \boldsymbol{y}' \iff \sum_{i=1}^{n} U_i(\boldsymbol{y}_i) \geq \sum_{i=1}^{n} U_i(\boldsymbol{y}'_i).$$

Proof of Theorem 1

Proof. One can easily check that (ii) implies (i). Let us show that (i) implies (ii). Consider a continuous index function $I : [0, \bar{y}]^n \to \mathbb{R}$ that satisfies monotonicity, separability, and the working age principle.

Step 1: find a function that is ordinally equivalent to I. To apply Debreu's Representation Theorem, generate a continuous and complete preordering $\succsim$ on $[0, \bar{y}]^n$ from I according to Euclidean distance $\geq$ on $\mathbb{R}$: for each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$,

$$\boldsymbol{y} \succsim \boldsymbol{y}' \iff I(\boldsymbol{y}) \geq I(\boldsymbol{y}'). \tag{4.17}$$

Note that the space $[0, \bar{y}]$ is connected.

First, we claim that the factors $1, \ldots, n$ are independent. For each $i \in N$, each $y_i, y_i' \in [0, \bar{y}]$, and each $\boldsymbol{y}_{-i}, \boldsymbol{y}'_{-i} \in [0, 1]^{n-1}$, by (4.17) and separability of I,

$$\begin{aligned} (y_i, \boldsymbol{y}_{-i}) \succsim (y_i', \boldsymbol{y}_{-i}) &\iff I(y_i, \boldsymbol{y}_{-i}) \geq I(y_i', \boldsymbol{y}_{-i}) \\ &\iff I(y_i, \boldsymbol{y}'_{-i}) \geq I(y_i', \boldsymbol{y}'_{-i}) \\ &\iff (y_i, \boldsymbol{y}'_{-i}) \succsim (y_i', \boldsymbol{y}'_{-i}). \end{aligned}$$

Second, we claim that all factors $i \in \{1, \ldots, n\}$ are essential. Take any $i \in \{1, \ldots, n\}$. Let $y_i, y_i' \in [0, \bar{y}]$ be such that $y_i > y_i'$. Then by monotonicity of I, for each $\boldsymbol{y}_{-i} \in [0, \bar{y}]^n$,

$$I(y_i, \boldsymbol{y}_{-i}) > I(y_i', \boldsymbol{y}_{-i}).$$

Hence

$$(y_i, \boldsymbol{y}_{-i}) \succsim (y_i', \boldsymbol{y}_{-i}).$$

Therefore, Debreu's Representation Theorem can be applied to $\succsim$ on $[0, \bar{y}]^n$. That is, for each $j \in \{1, \ldots, n\}$, there exists a continuous function $f_j : [0, \bar{y}] \to \mathbb{R}$ such that for each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$,

$$\boldsymbol{y} \succsim \boldsymbol{y}' \iff \sum_{j=1}^{n} f_j(y_j) \geq \sum_{j=1}^{n} f_j(y_j'). \tag{4.18}$$

Therefore,

$$I(\boldsymbol{y}) \geq I(\boldsymbol{y}') \iff \sum_{j=1}^{n} f_j(y_j) \geq \sum_{j=1}^{n} f_j(y_j'). \tag{4.19}$$

Step 2: $\exists f : [0, \bar{y}] \to \mathbb{R}$, $I(\boldsymbol{y}) \geq I(\boldsymbol{y}') \iff \sum_{j=1}^{n} f(y_j) \geq \sum_{j=1}^{n} f(y_j')$. Let us show that for each $j \in N$, $f_1 = f_j + f_1(0) - f_j(0)$. Take any $j \in N$. Suppose, by contradiction, that $f_1 \neq f_j + f_1(0) - f_j(0)$. Without loss of generality, consider the case with $f_1 > f_j + f_1(0) - f_j(0)$. Then, there exists $b \in [0, \bar{y}]$ such that $f_1(b) > f_j(b) + f_1(0) - f_j(0)$. Thus,

$$f_1(b) + f_j(0) > f_1(0) + f_j(b). \tag{4.20}$$

On the other hand, since

$$b + 0 = 0 + b \text{ and } a(b) + a(0) \geq a(0) + a(b),$$

by the working age principle

$$I(b, y_2, \ldots, \underbrace{0}_{j \text{ th}}, \ldots, y_n) \leq I(0, y_2 \ldots, \underbrace{b}_{j \text{ th}}, \ldots, y_n) \text{ for all } \boldsymbol{y}_{-1,j} \in [0, \bar{y}]^{n-2}.$$

Therefore,

$$f_1(b) + f_j(0) \leq f_1(0) + f_j(b),$$

a contradiction to equation (4.20). Therefore, $f_1 = f_j + f_1(0) - f_j(0)$. Then, by equation (4.19) of Step 1,

$$I(\boldsymbol{y}) \geq I(\boldsymbol{y}') \iff \sum_{j=1}^{n} f_1(y_j) \geq \sum_{j=1}^{n} f_1(y'_j).$$

For each $\alpha \geq 0$, let $f_\alpha : [0, \bar{y}] \to \mathbb{R}$ be such that for each $y_i \in [0, \bar{y}]$,

$$f_\alpha(y_i) = \begin{cases} y_i & \text{if } y_i < x, \\ (1+\alpha)y_i - \alpha x & \text{if } x \leq y_i \leq z, \\ y_i + \alpha(z - x) & \text{if } z < y_i. \end{cases}$$

Note that

$$I_\alpha(\boldsymbol{y}) = \sum_{j=1}^{n} f_\alpha(y_j).$$

Therefore, it suffices to show that there exists $\alpha \geq 0$ such that

$$\sum_{j=1}^{n} f(y_j) \geq \sum_{j=1}^{n} f(y_j) \iff \sum_{j=1}^{n} f_\alpha(y_j) \geq \sum_{j=1}^{n} f_\alpha(y_j).$$

Step 3: $\exists b_1, c_1 \in \mathbb{R}$, $\forall y_i \in [0, x)$, $f(y_i) = b_1 y_i + c_1$. Note that for each $y_i, y_j \in [0, x)$,

$$a(y_i) + a(y_j) = 2(z - x) = a\Big(\frac{y_i + y_j}{2}\Big) + a\Big(\frac{y_i + y_j}{2}\Big).$$

Then, since $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies the working age principle, for each $y_i, y_j \in [0, x)$,

$$f(y_i) + f(y_j) = 2f\Big(\frac{y_i + y_j}{2}, z\Big).$$

It in turn implies that for each $y_i, y_j \in [0, x)$,

$$\frac{1}{2}f(y_i) + \frac{1}{2}f(y_j) = f\Big(\frac{y_i + y_j}{2}\Big).$$

Then, by continuity of $f(\cdot)$ and Sierpinski Theorem, $f(\cdot)$ is convex and concave on $[0, x]$, that is, for each $y_i, y_j \in [0, x]$ and $\lambda \in [0, 1]$,

$$\lambda f(y_i) + (1 - \lambda) f(y_j) = f(\lambda y_i + (1 - \lambda) y_j). \tag{4.21}$$

Let

$$\begin{aligned} b_1 &= \frac{1}{x}(f(x) - f(0)), \\ c_1 &= f(0). \end{aligned}$$

We shall show that for each $y_i \in [0, x)$,

$$f(y_i) = b_1 y_i + c_1. \tag{4.22}$$

Take any $y_i \in [0, x)$. Then, since $y_i = \frac{y_i}{x} \cdot x + (1 - \frac{y_i}{x}) \cdot 0$, by (4.21),

$$\begin{aligned} f(y_i) &= \frac{y_i}{x} f(x) + (1 - \frac{y_i}{x}) f(0) \\ &= b_1 y_i + c_1. \end{aligned}$$

Therefore, (4.22) holds.

Step 4: $\exists b_2, c_2 \in \mathbb{R}$, $\forall y_i \in [x, z]$, $f(y_i) = b_2 y_i + c_2$. Note that for each $y_i, y_j \in [x, z]$,

$$a(y_i) + a(y_j) = y_i + y_j - 2x = a\Big(\frac{y_i + y_j}{2}\Big) + a\Big(\frac{y_i + y_j}{2}\Big).$$

Since $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies the working age principle, for each $y_i, y_j \in [x, z]$,

$$f(y_i) + f(y_j) = 2f\Big(\frac{y_i + y_j}{2}, z\Big),$$

that is,

$$\frac{1}{2}f(y_i) + \frac{1}{2}f(y_j) = f\Big(\frac{y_i + y_j}{2}\Big).$$

Then, by a similar argument to Step 1, there exists $b_2, c_2 \in \mathbb{R}$ such that for each $y_i \in [x, z]$, $f(y_i) = b_2 y_i + c_2$.

Step 5: $\exists b_3, c_3 \in \mathbb{R}$, $\forall y_i \in (z, \bar{y})$, $f(y_i) = b_3 y_i + c_3$. Note that for each $y_i, y_j \in (z, \bar{y}]$,

$$a(y_i) + a(y_j) = 0 = a\Big(\frac{y_i + y_j}{2}\Big) + a\Big(\frac{y_i + y_j}{2}\Big).$$

Since $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies the working age principle, for each $y_i, y_j \in (z, \bar{y}]$,

$$f(y_i) + f(y_j) = 2f\Big(\frac{y_i + y_j}{2}, z\Big),$$

that is,

$$\frac{1}{2}f(y_i) + \frac{1}{2}f(y_j) = f\Big(\frac{y_i + y_j}{2}\Big).$$

Then, by a similar argument to Step 1, there exists $b_3, c_3 \in \mathbb{R}$ such that for each $y_i \in (z, \bar{y}]$, $f(y_i) = b_3 y_i + c_3$.

Step 6: $0 < b_1 = b_3 \leq b_2$. By monotonicity, clearly $b_1, b_2, b_3 > 0$. Let us show that $b_1 = b_3$. Take any $y_i \in [0, x)$ and any $y_j \in (z, \bar{y}]$. Let $\epsilon > 0$ be such that $\epsilon < \min\{x - y_i, y_j - z\}$. Then, $y_i + \epsilon \in [0, x)$ and $y_j - \epsilon \in (z, \bar{y}]$. Moreover,

$$a(y_i) + a(y_j) = z - x = a(y_i + \epsilon) + a(y_j - \epsilon).$$

Therefore, by the working age principle,

$$f(y_i) + f(y_j) = f(y_i + \epsilon) + f(y_j - \epsilon),$$

that is,

$$b_1 y_i + c_1 + b_3 y_j + c_3 = b_1(y_i + \epsilon) + c_1 + b_3(y_j - \epsilon) + c_3.$$

It in turn implies that $b_1 = b_3$.

We next show that $b_3 \leq b_2$. Take any $y_i \in [x, z)$ and any $y_j \in (z, \bar{y}]$. Let $\epsilon > 0$ be such that $\epsilon < \min\{z - y_i, y_j - z\}$. Then, $y_i + \epsilon \in [x, z]$ and $y_j - \epsilon \in (z, \bar{y}]$. Moreover,

$$a(y_i) + a(y_j) = z - y_i > z - (y_i + \epsilon) = a(y_i + \epsilon) + a(y_j - \epsilon).$$

Therefore, by the working age principle,

$$f(y_i) + f(y_j) \leq f(y_i + \epsilon) + f(y_j - \epsilon),$$

that is,

$$b_2 y_i + c_2 + b_3 y_j + c_3 \leq b_2(y_i + \epsilon) + c_2 + b_3(y_j - \epsilon) + c_3.$$

It in turn implies that $b_3 \leq b_2$.

Step 7: $\boldsymbol{c_2 = -(b_2 - b_1)x + c_1}$, and $\boldsymbol{c_3 = (b_2 - b_1)(z - x) + c_1}$. By continuity of f at x, it follows that

$$b_1 x + c_1 = b_2 x + c_2.$$

Therefore,

$$c_2 = -(b_2 - b_1)x + c_1.$$

Similarly, by continuity of f at z, it follows that

$$b_2 z + c_2 = b_3 z + c_3.$$

Therefore, by $b_1 = b_3$,

$$c_3 = (b_2 - b_3)z + c_2 = (b_2 - b_1)z - (b_2 - b_1)x + c_1 = (b_2 - b_1)(z - x) + c_1.$$

Step 8: $\exists \alpha \geq 0$, $\sum_{j=1}^{n} f(y_j) \geq \sum_{j=1}^{n} f(x_j) \iff \sum_{j=1}^{n} f_\alpha(y_j) \geq \sum_{j=1}^{n} f_\alpha(x_j)$.

By Steps 1-5, for each $y_i \in [0, \bar{y}]$,

$$f(y_i) = \begin{cases} b_1 y_i + c_1 & \text{if } y_i < x, \\ b_2 y_i - (b_2 - b_1)x + c_1 & \text{if } x \leq y_i \leq z, \\ b_1 y_i + (b_2 - b_1)(z - x) + c_1 & \text{if } z < y_i. \end{cases}$$

Let

$$\alpha = \frac{b_2}{b_1} - 1.$$

Since $b_2 \geq b_1$, we have $\alpha \geq 0$. Then, for each $y_i \in [0, \bar{y}]$,

$$f(y_i) = b_1 f_\alpha(y_i) + c_1. \tag{4.23}$$

Therefore,

$$\sum_{j=1}^{n} f(y_j) \geq \sum_{j=1}^{n} f(x_j) \iff \sum_{j=1}^{n} f_\alpha(y_j) \geq \sum_{j=1}^{n} f_\alpha(x_j).$$

□

Proof of Corollary 1

Proof. One can easily check that (ii) implies (i). Note that if an index function I induces the same ordering as I_0, then I violates the strict working age principle. Therefore, by Theorem 1, (i) implies (ii). □

Proof of Corollary 2

Proof. One can easily check that (ii) implies (i). We show that (i) implies (ii). Since $I : [0, \bar{y}]^n \to \mathbb{R}$ satisfies continuity, monotonicity, separability, and normalization, there exists a continuous and strictly increasing function $f : [0, \bar{y}] \to \mathbb{R}$ such that

$$I(\boldsymbol{y}) = \frac{1}{\bar{y}} \cdot f^{-1}\Big(\frac{1}{n}\sum_{j=1}^{n} f(y_j)\Big).$$

Moreover, by the same argument as Proof of Theorem 1 (equation 4.23), there exist $\alpha, b > 0$ and $c \in \mathbb{R}$ such that

$$f(y_i) = bf_\alpha(y_i) + c.$$

Then, for each $s \in \mathbb{R}$,

$$f^{-1}(s) = f_\alpha^{-1}(\frac{s-c}{b}).$$

Therefore,

$$\begin{aligned}
I(\boldsymbol{y}) &= \frac{1}{\bar{y}} \cdot f^{-1}\Big(\frac{1}{n}\sum_{j=1}^{n} f(y_j)\Big) \\
&= \frac{1}{\bar{y}} \cdot f^{-1}\Big(\frac{1}{n}\sum_{j=1}^{n} (bf_\alpha(y_i) + c)\Big) \\
&= \frac{1}{\bar{y}} \cdot f_\alpha^{-1}\Big(\frac{\frac{1}{n}\sum_{j=1}^{n} (bf_\alpha(y_i) + c) - c}{b}\Big) \\
&= \frac{1}{\bar{y}} \cdot f_\alpha^{-1}\Big(I_\alpha(\boldsymbol{y})\Big).
\end{aligned}$$

□

Tightness of the axioms

- Let $I_1 : [0, \bar{y}]^n \to \mathbb{R}$ be an index function such that

$$I_1(\boldsymbol{y}) = \sum_{j=1}^{n}(z - x - a(y_j)) \text{ for all } \boldsymbol{y} \in [0, \bar{y}]^n.$$

Monotonicity: Let $\boldsymbol{y} = (0, 0, \ldots, 0)$ and $\boldsymbol{y'} = (x, x, \ldots, x)$. Then $\boldsymbol{y'} > \boldsymbol{y}$. By the definition of $a(\cdot)$, for each $j \in N$, $a(y_j) = a(y'_j)$. Therefore, $I_1(\boldsymbol{y}) = I_1(\boldsymbol{y'})$. Thus, I_1 violates monotonicity.

Separability: Consider any $\boldsymbol{y}, \boldsymbol{y}' \in \mathbb{R}^n_+$. Then, for each $i \in N$,

$$
\begin{aligned}
& \sum_{j \neq i} (z - x - a(y_j)) + z - x - a(y_i) \leq \sum_{j \neq i} (z - x - a(y'_j)) + z - x - a(y_i) \\
\iff & \sum_{j \neq i} (z - x - a(y_j)) + z - x - a(y'_i) \leq \sum_{j \neq i} (z - x - a(y'_j)) + z - x - a(y'_i)
\end{aligned}
$$

Therefore,

$$
I_1(y_j, \boldsymbol{y}_{-j}) \leq I_1(y_j, \boldsymbol{y'}_{-j}) \iff I_1(y'_j, \boldsymbol{y}_{-j}) \leq I_1(y'_j, \boldsymbol{y'}_{-j}).
$$

Thus, I_1 satisfies separability.

Working age principle: Consider any $\boldsymbol{y} \in \mathbb{R}^n_+$ and $y'_i, y'_j \in \mathbb{R}_+$. If $a(y_i) + a(y_j) = a(y'_i) + a(y'_j)$, then

$$
\sum_{k=1}^{n} (z - x - a(y_k)) = \sum_{k \neq i,j} (z - x - a(y'_j)) + z - x - a(y'_i) + z - x - a(y'_j).
$$

If $a(y_i) + a(y_j) < a(y'_i) + a(y'_j)$, then

$$
\sum_{k=1}^{n} (z - x - a(y_k)) > \sum_{k \neq i,j} (z - x - a(y'_j)) + z - x - a(y'_i) + z - x - a(y'_j).
$$

Thus, I_2 satisfies the strict working age principle.

- Let $I_2 : [0, \bar{y}]^n \to \mathbb{R}$ be an index function such that

$$
I_2(\boldsymbol{y}) = \sum_{j=1}^{n} y_j + \Big(\sum_{j=1}^{n} (z - x - a(y_j)) \Big)^2 \text{ for all } \boldsymbol{y} \in [0, \bar{y}]^n.
$$

Monotonicity and the working age principle: Obviously, I_2 satisfies monotonicity and the working age principle.

Separability: Let $g_2 : [0, \bar{y}] \to \mathbb{R}$ be such that for each $y_j \in [0, \bar{y}]$,

$$g_2(y_j) = \begin{cases} 0 & \text{if } y_i < x, \\ y_i - x & \text{if } x \leq y_i \leq z, \\ z - x & \text{if } z < y_i. \end{cases}$$

Then, $g_2(y_j) = z - x - a(y_j)$. Therefore, for each $\boldsymbol{y}, \boldsymbol{y}' \in [0, \bar{y}]^n$, we can compute that

$$\begin{aligned} & I_2(y_i, \boldsymbol{y}_{-i}) - I_2(y_i, \boldsymbol{y}'_{-i}) \\ & = \sum_{j \neq i} y_j - \sum_{j \neq i} y'_j + \Big(2g_2(y_i) + \sum_{j \neq i} g_2(y_j) + \sum_{j \neq i} g_2(y'_j)\Big)\Big(\sum_{j \neq i} g_2(y_j)) - \sum_{j \neq i} g_2(y'_j)\Big). \end{aligned} \tag{4.24}$$

Let $\varepsilon > 0$ be such that

$$\frac{1}{2}(x + 2\varepsilon^2 + \varepsilon) < x \text{ and } x + \varepsilon \leq z.$$

Fix some $i \in N$, and let

$$\boldsymbol{y}_{-i} = \Big(\frac{1}{2}(x + 2\varepsilon^2 + \varepsilon), \frac{1}{2}(x + 2\varepsilon^2 + \varepsilon), 0, \ldots, 0\Big) \in [0, \bar{y}]^{n-1},$$

$$\boldsymbol{y}'_{-i} = (x + \varepsilon, 0, \ldots, 0) \in [0, \bar{y}]^{n-1}.$$

Then, by equation (4.24), we have

$$I_2(0, \boldsymbol{y}_{-i}) - I_2(0, \boldsymbol{y}'_{-i}) = (x + 2\varepsilon^2 + \varepsilon) - (x + \varepsilon) + \Big(\varepsilon \cdot (-\varepsilon)\Big) = \varepsilon^2 > 0.$$

On the other hand, by equation (4.24), we have

$$I_2(x + \varepsilon, \boldsymbol{y}_{-i}) - I_2(x + \varepsilon, \boldsymbol{y}'_{-i}) = (x + 2\varepsilon^2 + \varepsilon) - (x + \varepsilon) + \Big((2\varepsilon + \varepsilon) \cdot (-\varepsilon)\Big) = -\varepsilon^2 < 0.$$

Thus, I_2 violates separability.

- Let $I_3 : [0,1]^n \to \mathbb{R}$ be the index function such that

$$I_3(\boldsymbol{y}) = \sum_{j=1}^{n} y_j^2 \text{ for all } \boldsymbol{y} \in [0, \bar{y}]^n.$$

Obviously I_3 satisfies monotonicity and separability.

Working age principle: Consider any $\boldsymbol{y}_{-i,j}$. Let $y_i = x$, $y_j = z$ and $y_i' = y_j = \frac{x+z}{2}$. Then, $y_i + y_j = y_i' + y_j'$ and $a(y_i) + a(y_j) = a(y_i') + a(y_j')$. But

$$\begin{aligned} I_3(\boldsymbol{y}) - I_3(y_i', y_j', \boldsymbol{y}_{-i,j}) &= x^2 + z^2 - \left(\frac{x+z}{2}\right)^2 \\ &= \frac{1}{2}(z-x)^2 > 0. \end{aligned}$$

Therefore, I_3 violate the working age principle.

These functions induce different orderings from I_α. The satisfaction and the violation of axioms by these functions are summarized by Table 7. It shows the independence of the axioms in our Theorem 1 and Corollary 1.

	MON	SEP	WAP	SWAP
I_1	−	+	+	+
I_2	+	−	+	+
I_3	+	+	−	−

Table 7: Tightness of axioms

References

Aczél, J. (1948) "On Mean Values," *Bulletin of the American Mathematical Society*, Vol.54, pp.392–400.

Aczél, J. (1987) *A Short Course on Functional Equations*, D. Reidel Publishing Company.

Alkire, S. and Foster, J. (2011a) "Counting and multidimensional poverty measurement," *Journal of Public Economics*, Vol. 95, pp.476–487.

Alkire, S. and Foster, J. (2011b) "Understandings and misunderstandings of multidimensional poverty measurement, " *Journal of Economic Inequality*, Vol.9 (2), pp.289–314.

Atkinson, A.B. (2003) "Multidimensional deprivation: Contrasting social welfare and counting approaches", *Journal of Economic Inequality*, Vol. 1(1), pp. 51–65.

Balinski, M. L. and Young, H. P. (1975) `The Quota Method of Apportionment," *American Mathematical Monthly*, Vol. 82, pp. 701–730.

Balinski, M. L. and Young, H. P. (1978) "Stability, Coalitions and Schisms in Proportional Representation Systems," *American Political Science Review*, Vol. 72, pp. 848–858.

Balinski, M. L. and Young, H. P. (1982) *Fair Representation: Meeting the Ideal of One Man, One Vote*. Yale University Press.

Bayardo, R. J., Ma, Y., and Srikant, R. (2007) "Scaling Up All Pairs Similarity Search," *Proceedings of the 16th International Conference on World Wide Web*, pp. 131–140.

Black, D. (1958) *The Theory of Committees and Elections.* Springer.

Black, D. (1976) "Partial justification of the Borda count," *Public Choice*, Vol. 28, pp. 1–15.

Borda, J.-C. de (1784) "Mémoire sur les Élections au Scrutin," *Histoire de l'Académie Royal des Sciences* 1981, pp. 657–664.

Chambers, C. P. (2008) "Consistent Representative Democracy," *Games and Economic Behavior*. Vol. 62, pp. 348–363.

Chambers, C. P. (2009) "An axiomatic Theory of Political Representation," *Journal of Economic Theory*. Vol. 144, pp. 375–389.

Chu, C.Y.C. (1997) "Age-Distribution Dynamics and Aging Indexes," *Demography*, Vol. 34(4), pp. 551–563.

Condorcet, M. de (1785) *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Reprinted by Chelfea Publishing Compeach (1972).

Coughlin, P. (1979) "A Direct Characterization of Black's First Borda Count," *Economics Letters*, Vol. 4, pp. 131–133.

Dalton, H. (1920) "The Measurement of the Inequality of Incomes," *Economic Journal*, Vol. 30, pp.348–361.

Debreu, G. (1959) "Topological Methods in Cardinal Utility Theory," In: K. J. Arrow, S. Karlin, and P. Suppes eds, *Mathematical Methods in the Social Sciences 1959: Proceedings of the First Stanford Symposium*, Stanford University Press, California, pp.16–26.

Desai, M. (1991) "Human Development: Concepts and Measurement," *European Economic Review*, Vol. 35, pp. 350–357.

Farkas, D. and Nitzan, S. (1979) "The Borda Rule and Pareto Stability: A Comment," *Econometrica*, Vol. 47, pp. 1305–1306.

Fishburn, P. C. and Gehrlein, W. V. (1976) "Borda's Rule, Positional Voting, and Condorcet's Simple Majority Principle," *Public Choice*, Vol. 28, pp. 79–88.

Foster, J.E. and Shorrocks, A.F. (1991) "Subgroup Consistent Poverty Indices," *Economterica*, Vol. 59(3), pp. 687–709.

Foster, J. E., Lopez-Calva, L. F., and Szekely, M. (2005) "Measuring the Distribution of Human Development: Methodology and an Application to Mexico," *Journal of Human Development*, Vol.6, pp.5–29.

Good, I. J. and Mittal, Y. (1987) "The Amalgamation and Geometry of Two-by-Two Contingency Tables," *Annals of Statistics*, Vol. 15, pp. 694–711.

Hardy, G. H., Littilewood, J. E., and Pólya, G. (1952) *Inequalities*, Cambridge Mathematical Library (Second edition).

Herrero, C., Martínez, R., and Villar, A. (2010) "Multidimensional Social Evaluation: an Application to the Measurement of Human Development," *Review of Income and Wealth*, Vol. 56, pp. 483–497.

Kawada, Y., Nakamura, Y., and Okamoto, N. (2017) "The Measurement of the Population Ageing," *mimeo, Keio Univetisty*.

Kawada, Y., Nakamura, Y., and Otani, S. (2018) "An Axiomatic Foundation of the Multiplicative Human Development Index," *Review of Income and Wealth*, forthcoming.

Keeney, R. L., and Raiffa H. (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley, New York.

Kemeny, J. G. (1959) "Mathematics Without Numbers," *Deadalus*, Vol. 88, pp. 577–591.

Kitao, S. (2015) "Fiscal Cost of Demographic Transition in Japan," *Journal of Economic Dynamics & Control*,Vol. 54, pp.37–58.

Maasoumi, E. and Lugo, M. A. (2007), "The information basis of multivariate poverty assessments", in (N. Kakwani and J. Silber, eds.), The Many Dimensions of Poverty, Palgrave-MacMillan.

Nguyen, H. V. and Bai, L. (2010) "Cosine Similarity Metric Learning for Face Verification," *Asian Conference on Computer Vision*, pp. 709–720.

Okamoto, N. and Sakai, T. (2013) "The Borda Rule and the Pairwise-majority Loser Revisited," unpublished manuscript, Keio University.

Rae, D. W. and Daudt, H. (1976) “The Ostrogorski Paradox: A Peculiarity of Compound Majority Decision," *European Journal of Political Research*, Vol. 4, pp. 391–398.

Saari, D. G. (1989) “A Dictionary for Voting Paradoxes," *Journal of Economic Theory*, Vol. 48, pp. 443–475.

Saari, D. G. (1990) “The Borda Dictionary,” *Social Choice and Welfare*, Vol. 7, pp. 279–317.

Saari, D. G. and Merlin, V. R. (2000) “A Geometric Examination of Kemeny’s Rule,” *Social Choice and Welfare*, Vol. 17, pp. 403–438.

Saari, D. G. (2006) “Which Is Better: the Condorcet and Borda Winner?," *Social Choice and Welfare*, Vol. 26, pp. 107–129.

Sagar, A. D., and Najam, A. (1998) “The Human Development Index: a Critical Reviews,” *Ecological Economics* Vol.25, pp. 249–264.

Sen, A. (1976) “Poverty: An Ordinal Approach to Measurement,” *Econometrica*, Vol. 44(2), pp. 219–231.

Sen, A. K. (1977) “Social choice theory: A re-examination," *Econometrica*, Vol. 45, pp. 53–89.

Sen, A. K. (1985) *Commodities and Capabilities*, North-Holland, Amsterdam.

Sen, A.K. (1990) “Development as capability expansion," In K. Griffin, & J. Knight (Eds.), Human development and the international development strategy for the 1990s. London: Macmillan.

Seth, S. (2013) “A Class of Distribution and Association Sensitive Multidimensional Welfare Indices,” *Journal of Economic Inequality* Vol.11, pp.133–162.

Shorrocks, A. F. (1980) “The Class of Additively Decomposable Inequality Measures,” *Econometrica* Vol.48, pp.613–625.

Singhal, A. (2001) “Modern Information Retrieval: A Brief Overview,” *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 24, pp. 35–43.

United Nations (2002) *World Population Ageing: 1950-2050.* United Nations, Population Division.

United Nations (2015) *United Nations Development Report 2015.* United Nations Development Programme.

United Nations (2017) *World Population Prospects 2017.* United Nations, Population Division.

Young, H. P. (1974) “An Axiomatization of Borda’s Rule," *Journal of Economic Theory*, Vol. 9, pp.43–52.

Young, H. P. (1975) “Social Choice Scoring Functions," *SIAM Journal of Applied Mathematics*, Vol. 28, pp. 824–838.

Young, H.P. (1988) “Condorcet’s Theory of Voting," *American Political Science Review*, Vol. 82, pp. 1231–1244.

Young, H.P. and Levenglick, A. (1978) “A Consistent Extension of Condorcet’s Election principle," *SIAM Journal on Applied Mathematics*, Vol. 35, pp. 285–300.

Zambrano, E. (2014) “An Axiomatization of the Human Development Index,” *Social Choice and Welfare*, Vol. 42, pp. 853–872.

Zhang, J. (2017) “The Evolution of China’s One-Child Policy and Its Effects on Family Outcomes," *Journal of Economic Perspectives*, Vol. 31(1), pp. 141–160.