Articles

New Goodness-of-Fit Measures for GARP and Critical Arcs

Takeshi Naitoh
Associate Professor / Shiga University

I Introduction

We consider a market analyst observing a finite dataset such as consumer purchase records. It is hypothesized that in such purchases, consumers make choices that provide the highest utility within a limited budget based on their own preferences. One of the themes of revealed preference theory is to test whether finite data are inconsistent with the utility maximization hypothesis. The study at the beginning of this line of research is Afriat [1]. Afriat's Theorem states that a given dataset is consistent with utility maximization if and only if it satisfies the generalized axiom of revealed preference (GARP). However, a single mistaken choice is enough to declare an entire dataset incompatible with rationality, even if all other choices could be explained as resulting from utility maximization. In practice, most choice datasets contain some revealed preference cycles that do not satisfy GARP.

In order to measure the degree of deviation of these violations, various and sundry goodness- of-fit indices for revealed preference conditions have been proposed in the literature (Lanier and Quah [10], Smeulders, Crama, and Spieksma [15]). The Afriat index (Afriat [2]) measures the percentage of a consumer's budget that is spent in violation of GARP. The Houtman-Maks index (Houtman and Maks [9]) indicates the largest percentage of observations that satisfy GARP. The minimum cost index (Dean and Martin [3]) is the minimum cost of removing revealed preference relations so that the remaining relations satisfy GARP. Shiozawa ([13]) proposed a goodness-of-fit measure (SCCI) using information obtained from strongly connected components. In this

* Faculty of Economics, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan

paper, we focus on the characteristics of strongly connected components and propose two new goodness-of-fit measures when there are linear budget constraints.

II Definitions and preliminaries

We introduce some notation used throughout the paper. There are n different types of goods in the market. The consumer has a budget b for consumption and a utility function U: $\mathbf{R}^n_+ \to \mathbf{R}$. We consider a market analyst observing a finite dataset $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$ over the time $t \in \{1, \ldots, T\}$, where $\boldsymbol{p}^t = (p^t_1, \ldots, p^t_n) \in \mathrm{R}^n_{++}$ is the positive price vector and $\boldsymbol{x}^t = (x^t_1, \ldots, x^t_n) \in \mathrm{R}^n_+ \setminus \{0\}$ is the consumer's demand bundle under prices $\boldsymbol{p}^t$ and the available budget. $b_t \in \mathbf{R}_+$. The inner product $\boldsymbol{p}^t \cdot \boldsymbol{x}^t = \Sigma^n_{i=1} \boldsymbol{p}^t_i \boldsymbol{x}^t_i$ represents total expenditure at time t. We assume $\boldsymbol{p}^t \cdot \boldsymbol{x}^t = b_t$. The dataset D is *rationalized* by a utility function U in the sense that for all $t \leqq T$, $\boldsymbol{x}^t$ maximizes U over $\{\boldsymbol{x} \mid \boldsymbol{p}^t \cdot \boldsymbol{x} \leqq \boldsymbol{p}^t \cdot \boldsymbol{x}^t\}$. The basic question raised by Afriat is whether the dataset is rationalized by a locally non-satiated utility function U.

A dataset D satisfies WARP if and only if, for each pair of distinct bundles $\boldsymbol{x}^i, \boldsymbol{x}^j$, $i, j \leqq T$ with $\boldsymbol{p}^i \cdot \boldsymbol{x}^i \geqq \boldsymbol{p}^i \cdot \boldsymbol{x}^j$, it is not the case that $\boldsymbol{p}^j \cdot \boldsymbol{x}^j \geqq \boldsymbol{p}^j \cdot \boldsymbol{x}^i$. We also say that the consumer's behavior satisfies Generalized Axiom of Revealed Preference (GARP) if $(\boldsymbol{p}^{t_1}, \boldsymbol{x}^{t_1}), (\boldsymbol{p}^{t_2}, \boldsymbol{x}^{t_2}), \ldots, (\boldsymbol{p}^{t_m}, \boldsymbol{x}^{t_m})$ satisfying $\boldsymbol{p}^{t_k} \cdot \boldsymbol{x}^{t_k} \geqq \boldsymbol{p}^{t_k} \cdot \boldsymbol{x}^{t_{k+1}}$ $(k = 1, \ldots, m-1)$ for all $t_1, \ldots, t_m \leqq T$, we have $\boldsymbol{p}^{t_m} \cdot \boldsymbol{x}^{t_1} \geqq \boldsymbol{p}^{t_m} \cdot \boldsymbol{x}^{t_m}$. A utility function U is said to *rationalize* the observed dataset D if $U(\boldsymbol{x}^t) \geqq U(\boldsymbol{x})$ for all $\boldsymbol{x}$ such that $\boldsymbol{p}^t \cdot \boldsymbol{x}^t \geqq \boldsymbol{p}^t \cdot \boldsymbol{x}$.

Theorem 2.1 (Afriat's Theorem [1], [19]): *The following four statements are equivalent:*
(a) *The dataset D can be rationalized by a locally non-satiated utility function U.*
(b) *The dataset D satisfies GARP.*
(c) *There is a positive solution* ϕ, λ *to the set of linear inequalities* $\lambda_j \leqq \phi_i + \lambda_i \boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i)$ *for all i, j.*
(d) *The dataset D can be rationalized by a continuous, concave, strictly monotone increasing utility function U.*

□

Note that a utility maximizer may violate GARP if the n types of goods are discrete, i.e., $x^t \in \mathbf{Z}^n_+ \setminus \{0\}$. S. Fujishige and Z. Yang [8] extended the theory of revealed preference to discrete models and established a discrete analogue of Afriat's theorem using a concept called tight budget demand set and the properties of strongly connected components of graphs. Polisson and Quah [11] and Forges and Iehè[6] also considered rationalizability for indivisible goods.

Let V and A be finite sets. A *graph* G is a pair (V, A) where V is the set of vertices and A is the set of arcs. We often denote it by $G = (V, A)$. A *directed graph* $G = (V, A)$ consists of a set V of vertices and a set A of arcs whose elements are ordered pairs of distinct vertices. If $a \in A$, $i, j \in V$, and $a = (i, j)$, then we say that *a joins i* to *j*. Also we call that i is the *tail* of a and j is the *head* of a. A *path* in $G = (V, A)$ is a sequence $P = (i_1, \ldots, i_l)$ of different vertices i_k $(k = 1, \ldots, l)$ such that $(i_k, i_{k+1}) \in A$ $(k = 1, \ldots, l-1)$. The end vertices of this path are i_1 and i_m, and the path is said to be an (i_1, i_l)-*path*. If P is an (i_1, i_l)-path in $G = (V, A)$ and $a \in A$ is an arc that joins i_l to i_1, then $C = (i_1, \ldots, i_l, i_1)$ is called a *cycle*.

A graph $G' = (W, B)$ is called a *subgraph of* $G = (V, A)$ if $W \subseteq V$ and $B \subseteq A$. For a vertex subset $W \subseteq V$, the subgraph $G[W]$ of G whose vertex set is W and whose arc set consists of the arcs of G joining vertices of W is called the subgraph of G induced by W. We denote by $G - a$ the graph obtained from $G = (V, A)$ by deleting the arc $a \in A$. Furthermore, if $B \subseteq A$, we denote by $G - B$ the graph obtained by deleting the arcs in B. A (sub)graph H is said to be *strongly connected* if for every two vertices i, j in the graph H there exists a path in H from i to j. A maximal strongly connected subgraph of a graph $G = (V, A)$ is called a *strongly connected component* of the graph G. G is decomposed into its strongly connected components $H_k = (V_k, A_k)$ $(k \in K)$ where $\{V_k \mid k \in K\}$ is a partition of V. An algorithm by Tarjan ([17]) finds a partition in linear time, $O(|V| + |E|)$.

We use the data matrix $D_T = (\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i))$ to construct a directed graph $G_D = (V_T, A)$, where $V_T = \{1, 2, \ldots, T\}$ is the set of vertices corresponding to the indices $1, 2, \ldots, T$ of the observations, and for $i, j \in V_T$ with $i \neq j$ the ordered pair $(i, j) \in A$ is an arc with a length $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i)$. A cycle is called a *negative length cycle* if the total length of the contained arcs in the cycle is strictly less than zero. For a directed graph $G_D = (V_T, A)$ we define the following directed subgraph $G_D^{\leqq 0} = (V_T, A^{\leqq 0})$ where $A^{\leqq 0}$ is the set of arcs with negative length, i.e., $A^{\leqq 0} = \{(i, j) \in A \mid \boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) \leqq 0\}$.

Proposition 2.2 ([8]): *The following five statements are equivalent:*
(a) *The data matrix D_T satisfies GARP.*
(b) *Every cycle C in the graph $G_D^{\leqq 0}$ satisfies $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$ for all $(i, j) \in C$.*
(c) *Every negative length cycle in the graph G_D contains at least one arc (i, j) such that $p^i(\boldsymbol{x}^j - \boldsymbol{x}^i) > 0$.*
(d) *Every cycle in the graph G_D that contains an arc of negative length must also contain an arc of positive length.*
(e) *Every strongly connected component $H_k = (V_k, A_k)$ of the graph $G_D^{\leqq 0}$ satisfies $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$ for all $(i, j) \in A_k$.* □

GARP is equivalent to what Afriat called cyclical consistency (Proposition 2.2 (b)). The cyclic consistency plays a fundamental role in the various literature on revealed preference (Dziewulski, Lanier, and Quah [5]). Algorithms for fast verification of GARP have been developed (Talla Nobibon, B. Smeulders, and F. C. R. Spieksma [16], etc.).

III Goodness-of-fit measures for GARP

For a given dataset, the revealed preference tests give results that either rationalizable or not. However, we are sometimes interested in the degree of these violations. A lot of goodness-of-fit measures for rationality have been proposed. In this section, we explain four goodness-of-fit measures.

Afriat [2] defines a partial efficiency index.

For a given real number e with $0 \leqq e \leqq 1$, $\boldsymbol{x}^t$ is directly revealed preferred to x at efficiency level e if $e\boldsymbol{p}^t \cdot \boldsymbol{x}^t \geqq \boldsymbol{p}^t \cdot x$. We say that the consumer's overall behavior satisfies e-Generalized Axiom of Revealed Preference (e-GARP) if $(\boldsymbol{p}^{t_1}, \boldsymbol{x}^{t_1})$, $(\boldsymbol{p}^{t_2}, \boldsymbol{x}^{t_2}), \ldots, (\boldsymbol{p}^{t_m}, \boldsymbol{x}^{t_m})$ satisfying $e\boldsymbol{p}^{t_k} \cdot \boldsymbol{x}^{t_k} \geqq \boldsymbol{p}^{t_k} \cdot \boldsymbol{x}^{t_{k+1}}$ $(k = 1, \ldots, m-1)$ for all $t_1, \ldots, t_m \leqq T$, we have $\boldsymbol{p}^{t_m} \cdot \boldsymbol{x}^{t_1} \geqq e\boldsymbol{p}^{t_m} \cdot \boldsymbol{x}^{t_m}$. If $e = 1$, this is the standard direct revealed preference relation. If $e = 0$, e-GARP is always satisfied. Hence there is some critical level e^* where the data just satisfy e-GARP.

Definition 3.1: (Afriat's efficiency index)
For a dataset $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$, *the Afriat's efficiency index* (AEI) *is defined as follows:*

$$\mathrm{AEI}(D) = \sup_{0 \leqq e \leqq 1} \left\{ e \,\middle|\, D \text{ satisfies } e\text{--GARP} \right\}.$$

□

The Houtman-Maks index reports the largest number of elements of subset of observations satisfying GARP. For a finite set X we denote its cardinality by $|X|$.

Definition 3.2: (*Houtman-Maks index*)
For a dataset $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$ *the Houtman-Maks index is defined as follows:*

$$\mathrm{HMI}(D) = \max_{X \subseteq \{1, \ldots, T\}} \left\{ \frac{|X|}{T} \,\middle|\, D \text{ satisfies GARP} \right\}.$$

□

Dean and Martin [3] proposed a goodness-of-fit measure based on Afriat's cyclical consistency.

Definition 3.3: (*Dean and Martin's minimum cost index*)
For a dataset $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$ *the minimum cost index* (MCI) *is defined as follows:*

$$\mathrm{MCI}(D) = \min \left\{ \frac{SA'}{\sum_{t=1}^{T} \boldsymbol{p}^t \cdot \boldsymbol{x}^t} \,\middle|\, A' \subseteq A^{\leqq 0} \text{ and } G' = (V_M, A^{\leqq 0} \setminus A') \text{ contains no directed cycle} \right\}$$

where $SA' = \sum_{(i,j) \in A'} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)$. □

On the other hand Shiozawa [13] paid attention to the fact that if there is a data which violates GARP then a relevant negative length arc is contained in a strongly connected component of the graph $G^{\leqq 0}$.

Definition 3.4: (*Shiozawa's strongly connected component index* [13])
For a dataset $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$ *the strongly connected component index* (SCCI) *is defined as follows:*

$$\mathrm{SCCI}(D) = \frac{\displaystyle\sum_{k \in K} \sum_{(i,j) \in A_k} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)}{\displaystyle\sum_{(i,j) \in A^{\leqq 0}} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)}$$

where $H_k = (V_k, A_k)$ $(k \in K)$ *is the strongly connected component decomposition of* $G^{\leqq 0}$. *If any* $(i, j) \in A^{\leqq 0}$ *satisfies* $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$, Index $1(D) = 0$. □

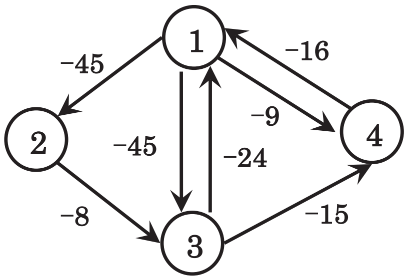
Example 3.5: *Let* $t = 1, 2, 3, 4$. *Suppose that the dataset* D_1 *is given by* $\{(p^t, x^t)\} \mid t \leqq 4\} = \{((9, 15), (40, 31)), ((10, 14), (60, 16)), ((12, 13), (55, 19)), ((13, 12), (44, 28))\}$ *and that the consumer has a budget of* $b_1 = 825$, $b_2 = 824$, $b_3 = 907$, $b_4 = 908$.

$$\begin{pmatrix} 9 & 15 \\ 10 & 14 \\ 12 & 13 \\ 13 & 12 \end{pmatrix} \begin{pmatrix} 40 & 60 & 55 & 44 \\ 31 & 16 & 19 & 28 \end{pmatrix} = \begin{pmatrix} 825 & 780 & 780 & 816 \\ 834 & 824 & 816 & 832 \\ 883 & 928 & 907 & 892 \\ 892 & 972 & 943 & 908 \end{pmatrix}.$$

Then

$$D_{1T} = \begin{pmatrix} 0 & -45 & -45 & -9 \\ 10 & 0 & -8 & 8 \\ -24 & 21 & 0 & -15 \\ -16 & 64 & 35 & 0 \end{pmatrix}$$

and for the directed subgraph $G_1^{\leqq 0} = (V_T, A_1^{\leqq 0})$, $A_1^{\leqq 0} = \{(1, 2), (1, 3), (1, 4), (2, 3), (3, 1), (3, 4), (4, 1)\}$.



$G_1^{\leqq 0}$ *contains cycles* $(1, 3, 1)$, $(1, 2, 3, 4, 1)$ *and so on. Hence the observed data is not consistent with GARP.* □

First, we consider AEI(D_1) for Example 3.5. If $e < \frac{883}{907}$, then $A^{\leqq 0} = \{(1, 2), (1, 3)\}$ and the graph $G^{\leqq 0}$ does not contain a cycle. Else if $e = \frac{883}{907}$, then $A^{\leqq 0} = \{(1, 2), (1, 3), (3, 1)\}$ and $G^{\leqq 0}$ contains the cycle (1, 3, 1). Hence

$$\mathrm{AEI}(D_1) = \frac{883}{907} \fallingdotseq 0.974.$$

Next, we consider HMI(D_1), MCI(D_1), and SCCI(D_1) for Example 3.5. Note that $G^{\leqq 0}$ contains cycles (1, 3, 1), (1, 4, 1), (1, 2, 3, 1), (1, 2, 3, 4, 1) and $G^{\leqq 0}$ is strongly connected component. Hence

$$\mathrm{HMI}(D_1) = \frac{|\{2, 3, 4\}|}{4} = 0.75.$$

$$\mathrm{MCI}(D_1) = \frac{24 + 16}{825 + 824 + 907 + 908} = \frac{40}{3464} \fallingdotseq 0.0116.$$

$$\mathrm{SCCI}(D_1) = \frac{45 + 45 + 9 + 8 + 24 + 15 + 16}{45 + 45 + 9 + 8 + 24 + 15 + 16}. = \frac{162}{162} = 1.$$

Note that SCCI always returns a value of 1 when $G^{\leqq 0}$ is a single strongly connected component. However, it seems to be a proper indicator when there are multiple strongly connected components.

Example 3.6: *Let* $t = 1, 2, 3, 4$. *Suppose that the dataset* D_3 *is given by*

$$\left\{(\boldsymbol{p}^t, \boldsymbol{x}^t) \,|\, t \leqq 4\right\} = \{((3, 4), (12, 16)), ((4, 3), (19, 8)), ((4, 4), (21, 5)), ((4, 5), (15, 10))\}$$

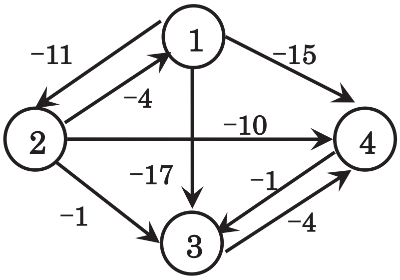
and that the consumer has a budget of $b_1 = b_2 = 100$, $b_3 = 104$, $b_4 = 110$.

$$\begin{pmatrix} 3 & 4 \\ 4 & 3 \\ 4 & 4 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 12 & 19 & 21 & 15 \\ 16 & 8 & 5 & 10 \end{pmatrix} = \begin{pmatrix} 100 & 89 & 83 & 85 \\ 96 & 100 & 99 & 90 \\ 112 & 108 & 104 & 100 \\ 128 & 116 & 109 & 110 \end{pmatrix}.$$

Then

$$D_{2T} = \begin{pmatrix} 0 & -11 & -17 & -15 \\ -4 & 0 & -1 & -10 \\ 8 & 4 & 0 & -4 \\ 18 & 6 & -1 & 0 \end{pmatrix}$$

and $A^{\leqq 0} = \{(1, 2), (1, 3), (1, 4), (2, 1), (2, 3), (2, 4), (3, 4), (4, 3)\}$. $G^{\leqq 0}$ *contains cycles* (1, 2, 1) *and* (3, 4, 3).



Strongly connected components of the graph $G^{\leqq 0}$ *are* $H_1 = (\{1, 2\}, \{(1, 2), (2, 1)\})$ *and* $H_2 = (\{3, 4\}, \{(3, 4), (4, 3)\})$. □

If $e < \frac{96}{100}$, then $A^{\leqq 0} = \{(1, 2), (1, 3), (1, 4), (2, 4)\}$ and the graph $G^{\leqq 0}$ does not contain a cycle. Else if $e = \frac{96}{100}$, then $A^{\leqq 0} = \{(1, 2), (1, 3), (1, 4), (2, 1), (2, 4)\}$ and $G^{\leqq 0}$ contains the cycle $(1, 2, 1)$. Hence

$$\mathrm{AEI}(D_2) = \frac{96}{100} = 0.96.$$

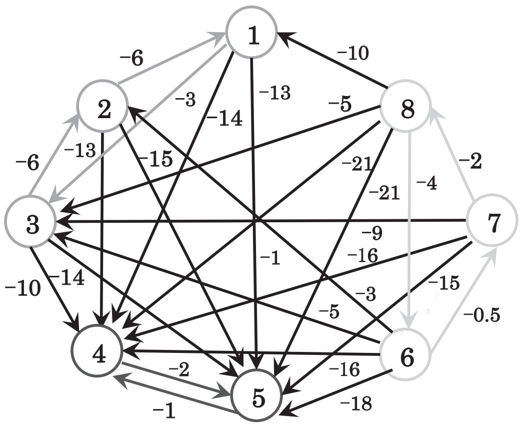
$$\mathrm{HMI}(D_2) = \frac{|\{1, 3\}|}{4} = 0.5.$$

$$\mathrm{MCI}(D_2) = \frac{4 + 1}{100 + 100 + 104 + 110} = \frac{5}{414} \fallingdotseq 0.0121.$$

$$\mathrm{SCCI}(D_2) = \frac{11 + 4 + 4 + 1}{11 + 17 + 15 + 4 + 1 + 10 + 4 + 1} = \frac{20}{63} \fallingdotseq 0.317.$$

Example 3.7: *Let* $t = 1, \ldots, 8$. *Suppose that the dataset* $D_3 = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t \leqq 8\}$ *is given by* $\{((3, 5, 5), (27, 24, 29)), ((4, 5, 4), (21, 34, 24)), ((4, 4, 5), (31, 28, 22)), ((4, 5, 5), (29, 25, 24)), ((3, 5, 6), (26, 27, 24)), ((4, 5, 6), (29, 27, 25)), ((3, 5, 7), (27.5, 30.5, 23)), ((4, 6, 5), (31, 28, 23))\}$ *and that the consumer has a budget of* $b_1 = 346$, $b_2 = 350$, $b_3 = 346$, $b_4 = 361$, $b_5 = 357$, $b_6 = 401$, $b_7 = 396$, $b_8 = 407$. *Corresponding data matrix* D_T *is*

$$D_{3T} = \begin{pmatrix} 0 & 7 & -3 & -14 & -13 & 1 & 4 & 2 \\ -6 & 0 & 2 & -13 & -15 & 1 & 4.5 & 6 \\ 3 & -6 & 0 & -10 & -14 & 3 & 1 & 5 \\ 12 & 13 & 13 & 0 & -2 & 15 & 16.5 & 18 \\ 18 & 20 & 8 & -1 & 0 & 15 & 16 & 14 \\ 1 & -3 & -5 & -16 & -18 & 0 & -0.5 & 1 \\ 8 & 5 & -9 & -16 & -15 & 1 & 0 & -2 \\ -10 & 1 & -5 & -21 & -21 & -4 & 1 & 0 \end{pmatrix}$$



$G_2^{\leqq 0}$ contains three strongly connected components. □

$$\mathrm{AEI}(D_3) = \frac{343}{346} \fallingdotseq 0.991.$$

$$\mathrm{HMI}(D_3) = \frac{|\{1,2,4,6,7\}|}{8} = 0.625.$$

$$\mathrm{MCI}(D_3) = \frac{0.5+1+3}{346+350+346+361+357+401+396+407} = \frac{4.5}{2964} \fallingdotseq 0.00152.$$

$$\mathrm{SCCI}(D_3) = \frac{(6+6+3)+(2+1)+(0.5+2+4)}{\sum_{(i,j)\in A^{\leqq 0}} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)} = \frac{24.5}{242.5} \fallingdotseq 0.101.$$

IV Critical arcs

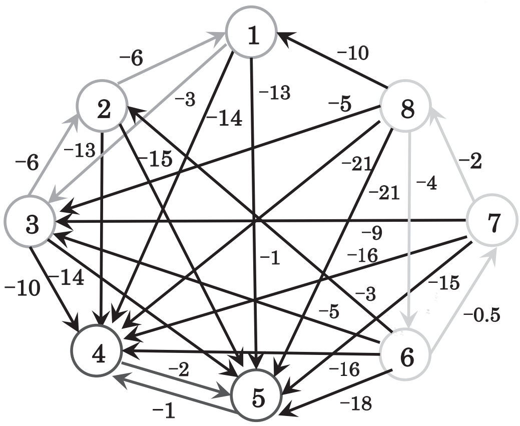
SCCI uses all negative length arcs of the strongly connected components. However, the information that some of the length could be improved to satisfy GARP seems to be lost. In this section, we propose a concept called *critical arcs* about the arcs to be removed so that the sum of the length is as small as possible under the condition that the remaining relationships satisfy GARP. First, examples are used to illustrate the procedures for identifying critical arcs.

We denote by $A_{SCC}(G)$ the set of arcs of the strongly connected components of G. Decompose $G_D^{\leqq 0}$ into strongly connected components. If any $(i,j) \in A_{SCC}(G_D^{\leqq 0})$ satisfies $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$, GARP is satisfied. If GARP is not satisfied, then there exists at least one arc with negative length contained in $(i,j) \in A_{SCC}(G_D^{\leqq 0})$.

Choose one arc $a_1 = (i_1, j_1) \in A_{SCC}(G_D^{\leqq 0})$ with the maximum length l_1, but if there is more than one arc with the length l_1, select the arc with the maximum value of $\frac{\boldsymbol{p}^i \cdot \boldsymbol{x}^j}{\boldsymbol{p}^i \cdot \boldsymbol{x}^i}$. If the values of $\frac{\boldsymbol{p}^i \cdot \boldsymbol{x}^j}{\boldsymbol{p}^i \cdot \boldsymbol{x}^i}$ are also equal, choose the arc with the smallest tail number i, and if the number i

is also equal, choose the arc with the smallest head number j. The arc $a_1 = (i_1, j_1)$ is a critical arc.

Example 4.1: *For Example 3.5, $A_{SCC}(G_{D_1}^{\leqq 0}) = A^{\leqq 0}$ and a critical arc is $(2, 3)$ with length $l_1 = -8$. For Example 3.6, $A_{SCC}(G_{D_2}^{\leqq 0}) = \{(1, 2), (2, 1), (3, 4), (4, 3)\}$ and a critical arc is $(4, 3)$ with length $l_1 = -1$. For Example 3.7*



$A_{SCC}(G_{D_3}^{\leqq 0}) = \{(1, 3), (3, 2), (2, 1), (4, 5), (5, 4), (6, 7), (7, 8), (8, 6)\}$ and a critical arc is $(6, 7)$ with length $l_1 = -0.5$. □

Example 4.2: *Let $t = 1, \ldots, 7$. Suppose that the dataset D_4 is given by*

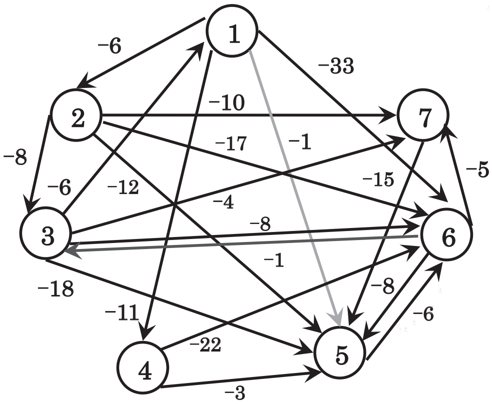
$$\begin{aligned}&\{(p^t, x^t) \mid t \leqq 7\}\\ &= \{((4, 5, 4), (30, 44, 34)), ((3, 5, 6), (38, 34, 37)), ((4, 4, 6), (44, 36, 31)), ((4, 5, 5), (39, 29, 41)),\\ &\quad ((4, 5, 6), (37, 43, 28)), ((3, 5, 7), (36, 27, 41)), ((4, 4, 7), (44, 38, 29))\}\end{aligned}$$

and that the consumer has a budget of $b_1 = 476$, $b_2 = b_3 = b_4 = 506$, $b_5 = 531$, $b_6 = 530$, $b_7 = 531$.

$$\begin{pmatrix} 4 & 5 & 4 \\ 3 & 5 & 6 \\ 4 & 4 & 6 \\ 4 & 5 & 5 \\ 4 & 5 & 6 \\ 3 & 5 & 7 \\ 4 & 4 & 7 \end{pmatrix} \begin{pmatrix} 30 & 38 & 44 & 39 & 37 & 36 & 44 \\ 44 & 34 & 36 & 29 & 43 & 27 & 38 \\ 34 & 37 & 31 & 41 & 28 & 41 & 29 \end{pmatrix} = \begin{pmatrix} 476 & 470 & 480 & 465 & 475 & 443 & 482 \\ 514 & 506 & 498 & 508 & 494 & 489 & 496 \\ 500 & 510 & 506 & 518 & 488 & 498 & 502 \\ 510 & 507 & 511 & 506 & 503 & 484 & 511 \\ 544 & 544 & 542 & 547 & 531 & 525 & 540 \\ 548 & 543 & 529 & 549 & 522 & 530 & 525 \\ 534 & 547 & 537 & 559 & 516 & 539 & 531 \end{pmatrix}.$$

Hence the corresponding data matrix is

$$D_{4T} = \begin{pmatrix} 0 & -6 & 4 & -11 & -1 & -33 & 6 \\ 8 & 0 & -8 & 2 & -12 & -17 & -10 \\ -6 & 4 & 0 & 12 & -18 & -8 & -4 \\ 4 & 1 & 5 & 0 & -3 & -22 & 5 \\ 13 & 13 & 11 & 16 & 0 & -6 & 9 \\ 18 & 13 & -1 & 19 & -8 & 0 & -5 \\ 3 & 16 & 6 & 28 & -15 & 8 & 0 \end{pmatrix}.$$



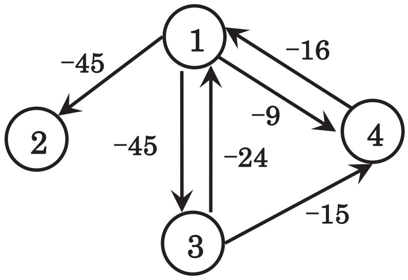
Since $G_4^{\leqq 0}$ *is strongly connected component,* $A_{SCC}(G_{D_4}^{\leqq 0}) = A^{\leqq 0}$*. The maximum length* l_1 *of each arc is* -1*, which corresponds to the arcs* $(1, 5)$ *and* $(6, 3)$*.*

$$\frac{\boldsymbol{p}^1 \cdot \boldsymbol{x}^5}{\boldsymbol{p}^1 \cdot \boldsymbol{x}^1} = \frac{475}{476} \fallingdotseq 0.9979, \quad \frac{\boldsymbol{p}^6 \cdot \boldsymbol{x}^3}{\boldsymbol{p}^6 \cdot \boldsymbol{x}^6} = \frac{529}{530} \fallingdotseq 0.9981.$$

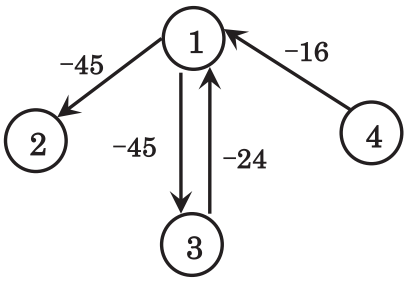
Hence a critical arc is $(6, 3)$. □

Next, decompose $G_D^{\leqq 0} - a_1$ into strongly connected components. If $G^{\leqq 0} - a_1$ satisfies GARP, then the critical arc of $G^{\leqq 0}$ is $a_1 = (i_1, j_1)$. Else choose one arc $a_2 = (i_2, j_2) \in A_{SCC}(G_D^{\leqq 0} - a_1)$ based on the same rules as when we chose a_1 . Repeat these procedures until $G^{\leqq 0} - \bigcup_{h=1}^{n} \{a_h\}$ satisfies GARP. Then critical arcs of $G_D^{\leqq 0}$ is $\{a_1, \ldots, a_n\}$.

Example 4.3: *For Example 3.5,* $a_1 = (i_1, j_1) = (2, 3)$ *with length* $l_1 = -8$. $A_{SCC}(G_{D_1}^{\leqq 0} - a_1) = \{(1, 3), (1, 4), (3, 1), (3, 4), (4, 1)\}$.



$a_2 = (1, 4)$ *with length* $l_2 = -9$.
$a_3 = (3, 4)$ *with length* $l_3 = -15$.
$A_{SCC}(G_{D_1}^{\leqq 0} - \{a_1, a_2, a_3\}) = \{(1, 3), (3, 1)\}$.



$a_4 = (3, 1)$ *with length* $l_4 = -24$.
Since $A_{SCC}(G_1^{\leqq 0} - \{a_1, a_2, a_3, a_4\}) = \varnothing$, *i.e.,* $G_1^{\leqq 0} - \{a_1, a_2, a_3, a_4\}$ *satisfies GARP. Hence critical arcs of* $G_1^{\leqq 0}$ *is* $\{(2, 3), (1, 4), (3, 4), (3, 1)\}$. □

Hereafter, the data matrix D_T is assumed to be non-negative, i.e., it has at least one negative element.
Algorithm
$G := G_D^{\leqq 0}$.

1. Decompose G into strongly connected components. If any $(i, j) \in A_{SCC}(G)$ satisfies $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$, GARP is satisfied, and the algorithm terminates.
2. Set $n := 1$ and $l := 0$, and repeat the following loop.

Choose one arc $a_n = (i_n, j_n) \in A_{SCC}(G)$ with the maximum length l_n, but if there is more than one arc with the length l_n, select the arc with the maximum value of $\frac{p^i \cdot x^j}{p^i \cdot x^i}$. If the values of $\frac{p^i \cdot x^j}{p^i \cdot x^i}$ are also equal, choose the arc with the smallest tail number i, and if the number i is also equal, choose the arc with the smallest head number j.

$G := G - a_n$
$l := l - l_n$

The loop terminates when any $(i, j) \in A_{SCC}(G)$ satisfies $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$.

Otherwise, $n := n + 1$ and return to the beginning.

V Our goodness-of-fit measures for GARP

In this section, we introduce two goodness-of-fit measures for GARP that use information on the length of the critical arcs. Denote by L the sum of absolute lengths of critical arcs. We focus on L as the numerator. This value L can be obtained using the algorithm in the previous section.

The first index we propose uses the sum of the lengths of the arcs of the strongly connected component of $G^{\leqq 0}$ as the denominator.

Definition 5.1: *For a dataset* $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$, *if some* $(i, j) \in A^{\leqq 0}$ *satisfies* $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) < 0$, *Index 1 is defined as follows:*

$$\text{Index } 1(D) = \frac{L}{\sum_{(i,j)\in A^{\leqq 0}} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)}.$$

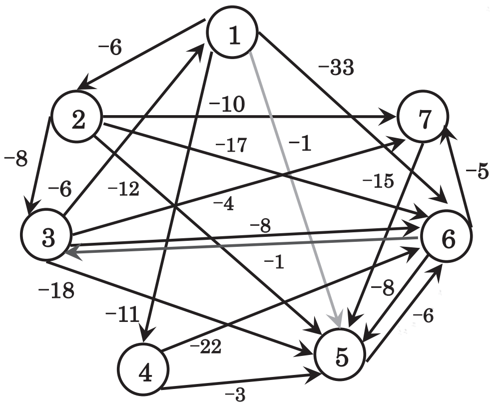
If any $(i, j) \in A^{\leqq 0}$ *satisfies* $\boldsymbol{p}^i \cdot (\boldsymbol{x}^j - \boldsymbol{x}^i) = 0$, *we define* Index $1(D) = 0$. □

Note that the denominator of Index 1 is equal to the numerator of SCCI.

Example 5.2: *For Example 3.5,*

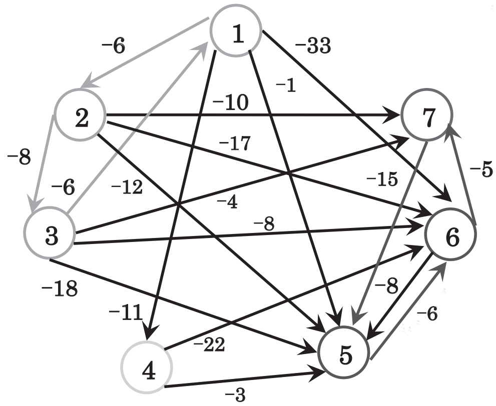
$$\text{Index } 1(D_1) = \frac{8 + 9 + 15 + 24}{45 + 45 + 9 + 8 + 24 + 15 + 16} = \frac{28}{81} \fallingdotseq 0.346.$$

Example 5.3: *For example 4.2,* $A_{SCC}(G^{\leqq 0}_{D_1}) = A^{\leqq 0}$.

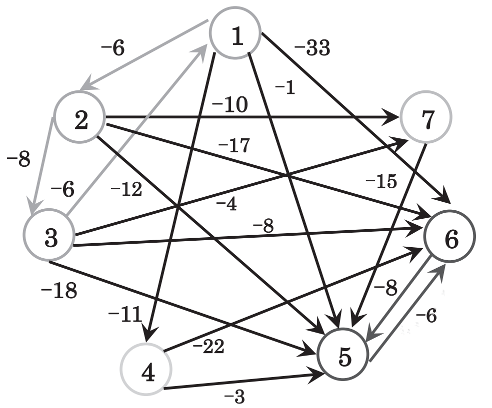


□

$a_1 = (6, 3)$ *with length* $l_1 = -1$.



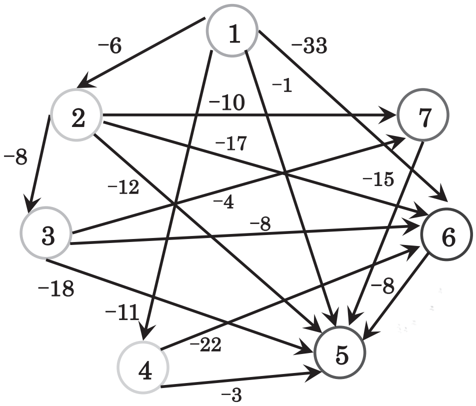
$A_{SCC}(G_{D_4}^{\leqq 0} - a_1) = \{(1,2), (2,3), (3,1), (5,6), (6,5), (6,7), (7,5)\}$. $a_2 = (6, 7)$ *with length* $l_2 = -5$.



$A_{SCC}(G_{D_4}^{\leqq 0} - \{a_1, a_2\}) = \{(1,2), (2,3), (3,1), (5,6), (6,5), (7,5)\}$. *There are three arcs of length* $-6 : (1,2), (3,1), (5,6)$.

$$\frac{\boldsymbol{p}^1 \cdot \boldsymbol{x}^2}{\boldsymbol{p}^1 \cdot \boldsymbol{x}^1} = \frac{470}{476} \fallingdotseq 0.987, \quad \frac{\boldsymbol{p}^3 \cdot \boldsymbol{x}^1}{\boldsymbol{p}^3 \cdot \boldsymbol{x}^3} = \frac{500}{506} \fallingdotseq 0.988, \quad \frac{\boldsymbol{p}^5 \cdot \boldsymbol{x}^6}{\boldsymbol{p}^5 \cdot \boldsymbol{x}^5} = \frac{525}{530} \fallingdotseq 0.989.$$

Hence $a_3 = (5, 6)$. $A_{SCC}(G_{D_4}^{\leqq 0} - \{a_1, a_2, a_3\}) = \{(1,2), (2,3), (3,1), (6,5), (7,5)\}$. $a_4 = (3, 1)$.



Since $A_{SCC}\,(G_4^{\leqq 0} - \{a_1, a_2, a_3, a_4\}) = \varnothing$, *each arc contained in* $A_{SCC}(G_1^{\leqq 0} - \{a_1, a_2, a_3, a_4\})$ *has length* 0. *Critical arcs of* $G_4^{\leqq 0}$ is $\{(6, 3), (6, 7), (5, 6), (3, 1)\}$.

$$\sum_{(i,j)\in A^{\leqq 0}} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j) = 194.$$

$$\text{Index } 1(D_4) = \frac{1+5+6+6}{194} = \frac{18}{194} \fallingdotseq 0.093.$$

MCI can be considered to effectively use information on the length of arcs, that need to be minimally modified to satisfy GARP. However, MCI is not polynomial time. The second index has the same numerator as Index 1 and the same denominator as MCI.

Definition 5.4: *For a dataset* $D = \{(\boldsymbol{p}^t, \boldsymbol{x}^t) \mid t = 1, \ldots, T\}$, *Index 2 is defined as follows:*

$$\text{Index } 2(D) = \frac{L}{\sum_{t\in T} \boldsymbol{p}^t \cdot \boldsymbol{x}^t}.$$

□

For Example 3.5,

$$\text{Index } 2(D_1) = \frac{8+9+15+24}{825+824+907+908} \fallingdotseq 0.016$$

For Example 4.2,

$$\text{Index } 2(D_4) = \frac{1+5+6+6}{476+506+506+506+531+530+531} = \frac{18}{3586} \fallingdotseq 0.00502.$$

	Index 1(D)	Index 2(D)	MCI(D)	SCCI(D)	1−AEI(D)	1−HMI(D)
Example3.5	0.346	0.016	0.012	1	0.026	0.25
Example3.6	0.25	0.012	0.012	0.317	0.04	0.5
Example3.7	0.019	0.0015	0.0015	0.101	0.009	0.375
Example4.2	0.093	0.0050	0.0036	1	0.012	0.286

VI Conclusion

SCCI uses all negative length arcs of the strongly connected components in the calculation of the numerator. By using the length of critical arcs instead, Index 1 may be more suitable than SCCI for cases with fewer strongly connected components of $G^{\leqq 0}$. Index 2 may be a relatively efficient way to obtain an approximate solution for MCI. However, it is a future issue that the ordering rule for arcs of the same length in the procedure for seeking a critical arc.

[Acknowledgments]

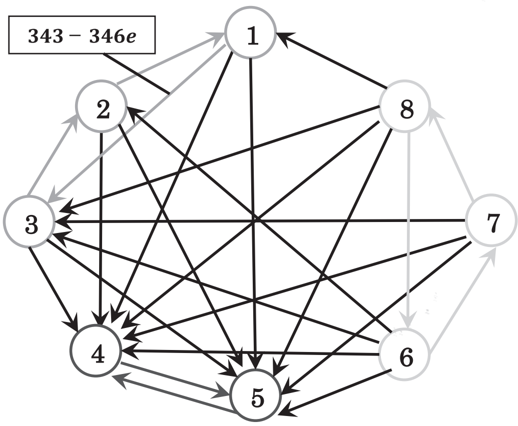
The author would like to express sincere gratitude to Satoru Fujishige for introducing this research theme and for giving insightful advice. The author also thanks the anonymous referee for helpful suggestions.

Appendix

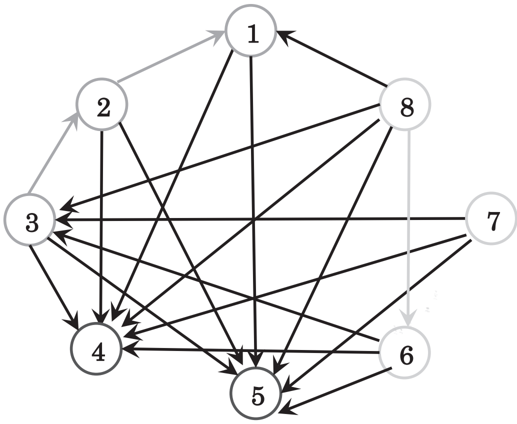
For example 3.7, we show AEI(D_3), HMI(D_3), MCI(D_3), Index 1(D_3), and Index 2(D_3).

For a given e its corresponding data matrix is

$$\begin{pmatrix}
0 & 353-346e & 343-346e & 332-346e & 333-346e & 347-346e & 350-346e & 348-346e \\
344-350e & 0 & 352-350e & 337-350e & 335-350e & 351-350e & 354.5-350e & 356-350e \\
349-346e & 340-346e & 0 & 336-346e & 332-346e & 349-346e & 347-346e & 351-346e \\
373-361e & 374-361e & 374-361e & 0 & 359-361e & 376-361e & 377.5-361e & 379-361e \\
375-357e & 377-357e & 365-357e & 356-357e & 0 & 372-357e & 373-357e & 371-357e \\
402-401e & 398-401e & 396-401e & 385-401e & 383-401e & 0 & 400.5-401e & 402-401e \\
404-396e & 401-396e & 387-396e & 380-396e & 381-396e & 397-396e & 0 & 394-396e \\
397-407e & 408-407e & 402-407e & 386-407e & 386-407e & 403-407e & 408-407e & 0
\end{pmatrix}$$



If $e < \frac{343}{346}$, then $G^{\leqq 0}_{D_3}$ is acyclic.



Else if $e = \frac{343}{346}$, then $G^{\leqq 0}$ contains the arc (1,3) with length 0 and the cycle (1, 3, 2, 1). Hence

$$\mathrm{AEI}(D_3) = \frac{343}{346} \fallingdotseq 0.991$$

Note that for Index HMI, $G^{\leqq 0}$ $[V \setminus t]$ $(t \in 1, 2, 3)$ contains two cycle ((4, 5, 4) and (6, 7, 8, 6)), $G^{\leqq 0}$ $[V \setminus t]$ $(t \in 4, 5)$ contains two cycle ((1, 3, 2, 1) and (6, 7, 8, 6)), and $G^{\leqq 0}$ $[V \setminus t]$ $(t \in 6, 7, 8)$ contains two cycle ((1, 3, 2, 1) and (4, 5, 4)). Therefore

$$\mathrm{HMI}(D_3) = \frac{|\{1, 2, 4, 6, 7\}|}{8} = 0.625.$$

If we delete {(6, 7), (5, 4), (1, 3)}, then $G^{\leqq 0}$ does not contain a cycle. Hence

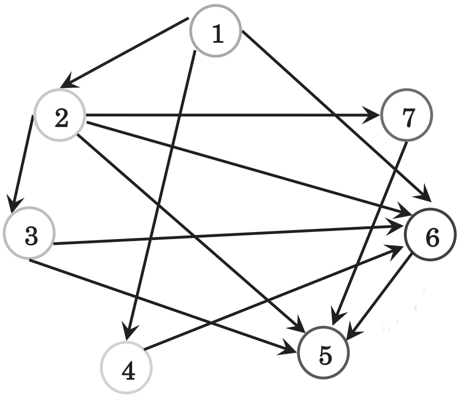
$$\mathrm{MCI}(D_3) = \frac{0.5 + 1 + 3}{346 + 350 + 346 + 361 + 357 + 401 + 396 + 407} \fallingdotseq 0.00152.$$

$$\mathrm{SCCI}(D_3) = \frac{(6 + 6 + 3) + (2 + 1) + (0.5 + 2 + 4)}{\displaystyle\sum_{(i,j) \in A^{\leqq 0}} \boldsymbol{p}^i \cdot (\boldsymbol{x}^i - \boldsymbol{x}^j)} = \frac{24.5}{242.5} \fallingdotseq 0.101.$$

$$\text{Index } 1(D_3) = \frac{0.5 + 1 + 3}{242.5} \fallingdotseq 0.019.$$

Moreover Index $2(D_3) = \mathrm{MCI}(D_3)$.

For Example 4.2 (the dataset D_4), we show $\mathrm{AEI}(D_4)$, $\mathrm{HMI}(D_4)$, $\mathrm{MCI}(D_4)$, and Index $2(D_4)$. If $e < \frac{500}{506}$, then $G_{D_4}^{\leqq 0}$ is acyclic.



$A^{\leqq 0} = \{(1, 2), (1, 4), (1, 6), (2, 3), (2, 5), (2, 6), (2, 7), (3, 5), (3, 6), (4, 6), (6, 5), (7, 5)\}$

Else if $e = \frac{500}{506}$, then $G^{\leqq 0}$ contains the cycle (1, 2, 3, 1). Hence

$$\mathrm{AEI}(D_4) = \frac{500}{506} \fallingdotseq 0.988.$$

Note that for Index HMI, $G^{\leqq 0}[V \setminus t]$ ($t \in 1, 2, 3, 4$) contains a cycle (5, 6, 7, 5) and $G^{\leqq 0}[V \setminus t]$ ($t \in 5, 6, 7$) contains a cycle (1, 2, 3, 1). Therefore

$$\mathrm{HMI}(D_4) = \frac{|\{1, 2, 4, 5, 7\}|}{7} \fallingdotseq 0.714.$$

If we delete $\{(1, 2), (5, 6), (6, 3)\}$, then $G^{\leqq 0}$ does not contain a cycle. Hence

$$\mathrm{MCI}(D_4) = \frac{6 + 6 + 1}{476 + 506 + 506 + 506 + 531 + 530 + 531} = \frac{13}{3586} \fallingdotseq 0.00363.$$

References

[1] S. N. Afriat: The Construction of Utility Function from Expenditure Data. *International Economic Review* 8(1) (1967) 67-77.

[2] S. N. Afriat: On a system of inequalities in demand analysis: An extension of the Classical Method. *International Economic Review* 14 (2) (1973) 460-472.

[3] M. Dean and D. Martin: Measuring Rationality with the Minimum Cost of Revealed Preference Violations. *Review of Economics and Statistics* 98(3) (2016) 524-534.

[4] T. Demuynck and J. Rehbeck: Computing revealed preference goodness-of-fitmeasures with integer programming. *Economic Theory* 76 (2023) 1175-1195.

[5] P Dziewulski, J. Lanier, and J. K. H. Quah: Revealed preference and revealed preference cycles: A survey. *Journal of Mathematical Economics* 113 (2024) 1-10.

[6] F. Forges and V. Iehlè Afriat's theorem for indivisible goods *Journal of Mathematical Economics* 54 (2014) 1-6.

[7] S. Fujishige: *Submodular Functions and Optimization*: Second Edition (North-Holland, Amsterdam, 2005).

[8] S. Fujishige and Z. Yang: On revealed preference and indivisibilities. *Modern Economy* 3(6) (2012) 752-758.

[9] M. Houtman and J. A. H. Maks: Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden* 6 (19) (1985) 89-104.

[10] J. Lanier and J. K.-H. Quah: Goodness-of-fit and utility estimation: what's possible and what's not. https://arxiv.org/abs/2405.08464 arXiv (2024) preprint

[11] M. Polisson and J. K.-H. Quah: Revealed Preference in a Discrete Consumption Space *American Economic Journal: Microeconomics* 5(1) (2013) 28-34.

[12] K. Shiozawa: Note on the goodness-of-fit measure for GARP; NP-hardness of minimum cost index. *Economics Bulletin* 3(4) (2015) 2455-2461.

[13] K. Shiozawa: Revealed preference test and shortest path problem; graph theoretic structure of the rationalizability test. *Journal of Mathematical Economics* 67 (2016) 38-48.

[14] B. Smeulders, F. C. R. Spieksma, L. Cherchye, and B. De Rock: Goodness-of-fit measures for Revealed Preference Tests: Complexity Results and Algorithms. *ACM Transactions on Economics and Computation archive* 2(1), Article 3 (2014).

[15] B. Smeulders, Y. Crama, and F. C. R. Spieksma: Revealed preference theory: An algorithmic outlook. *European Journal of Operational Research* 272(3) (2019) 803-815.

[16] F. Talla Nobibon, B. Smeulders, and F. C. R. Spieksma: A note on testing axioms of revealed preference. *Journal of Optimization Theory and Applications* 166(3) (2015) 1063- 1070.

[17] R. Tarjan: Depth-first search and linear graph algorithms. *SIAM Journal on Computing* 1(2) (1972) 146-160.

[18] H. R. Varian: Goodness-of-fit in optimizing models. *Journal of Econometrics* 46(1) (1990) 125-140.

[19] H. R. Varian: *Microeconomic analysis* : Third Edition (W. W. Norton & Company, New York, 1992).

New Goodness-of-Fit Measures for GARP and Critical Arcs

Takeshi Naitoh

It is a well-known result in revealed preference theory that a dataset is rationalizable if and only if the axiom called GARP is satisfied. We propose the concept of critical arcs, referring to the arcs to be removed from a graph representing a given dataset. These are collections of arcs whose total absolute length is as small as possible under the condition that the remaining relationships satisfy GARP after removal. We also propose two goodness-of-fit measures using critical arcs.

Key words: revealed preference, goodness-of-fit measures, strongly connected components