日本オペレーションズ・リサーチ学会
2021年 春季研究発表会

Measuring Olympic achievements in DEA: a data fitting technique subject to downside-deviation restrictions

01206313　東京理科大学　関谷 和之　SEKITANI Kazuyuki
05000602　東京理科大学　*趙　宇　ZHAO Yu

1. Introduction

Japan has become an Olympic setting since the Summer Olympic Games were first staged in Asia in 1964. After Tokyo was selected as the host city of the 2020 Summer Olympics, the Japanese Olympic Committee (JOC) set a target of a record-high 30 gold medals. However, due to the developing global situation in light of the COVID-19 pandemic, the International Olympic Committee (IOC) and JOC announced that the Tokyo 2020 Summer Olympic games are rescheduled to a date beyond 2020 but not later than summer 2021. As the upcoming Olympic Games gets closer, interest in predicting and measuring the Olympic achievements is rising. To investigate and predict the performance of the upcoming Summer Olympics, the proposed model of this study incorporates a data fitting technique of medal prediction using ordinary least squares regression (OLS) in input multiplier restrictions of the conventional DEA model.

2. Downside-deviation least squares

Suppose that there are n DMUs $(j = 1, \ldots, n)$ to be evaluated. Three output variables are considered: the number of gold (y_{1j}), silver (y_{2j}), and bronze medals (y_{3j}) achieved in the previous Olympic games. Regarding the inputs, we adopt the GDP (x_{1j}) and the population (x_{2j}) recorded in the latest year. The DEA technology converts an input vector $\boldsymbol{x}_j = (x_{1j}, x_{2j}, x_{3j})^\top$ to an output vector $\boldsymbol{y}_j = (y_{1j}, y_{2j}, y_{3j})^\top$, thereby satisfying $\boldsymbol{x}_j \geq \boldsymbol{0}, \boldsymbol{x}_j \neq \boldsymbol{0}$ and $\boldsymbol{y}_j \geq \boldsymbol{0}, \boldsymbol{y}_j \neq \boldsymbol{0}$ for all $j = 1, ..., n$.

Let $\boldsymbol{e}$ be a column vector with all elements unity. The OLS estimator for predicting the total number of medals $\boldsymbol{e}^\top \boldsymbol{y}_j$ of DMU$_j$ $(j \in \{1, ..., n\})$ is formulated as the following optimization problem:

(*OLS: Ordinary least squares*)

$$\begin{array}{ll} \min & \displaystyle\sum_{j=1}^{n} d_j^2 \\ \text{s.t.} & d_j = \boldsymbol{v}^\top \boldsymbol{x}_j + \nu - \boldsymbol{e}^\top \boldsymbol{y}_j \quad (j = 1, \ldots, n). \end{array}$$

Here, the estimation $(\boldsymbol{v}^*, \nu^*)$ presents the hyperplane passing the barycenter of n DMUs, $\frac{1}{n}\left(\sum_{j=1}^{n} \boldsymbol{x}_j, \sum_{j=1}^{n} \boldsymbol{e}^\top \boldsymbol{y}_j\right)$.

Since an upper bound $\boldsymbol{v}^\top \boldsymbol{x}_k + \nu$ of $\boldsymbol{e}^\top \boldsymbol{y}_k$ implies $\boldsymbol{v}^\top \boldsymbol{x}_k + \nu \geq \boldsymbol{e}^\top \boldsymbol{y}_k$, we add n nonnegative constraints, $d_1 \geq 0, \ldots, d_n \geq 0$, into the above OLS problem, which is formulated as

(*DLS: Downside-deviation least squares*)

$$\begin{array}{ll} \min & \displaystyle\sum_{j=1}^{n} d_j^2 \\ \text{s.t.} & 0 \leq d_j = \boldsymbol{v}^\top \boldsymbol{x}_j + \nu - \boldsymbol{e}^\top \boldsymbol{y}_j \quad (j = 1, \ldots, n) \\ & \boldsymbol{v} \geq \boldsymbol{0}. \end{array}$$

Let $(\hat{\boldsymbol{d}}, \hat{\boldsymbol{v}}, \hat{\nu})$ be an optimal solution to the DLS problem. Thus, $\hat{\boldsymbol{v}}^\top \boldsymbol{x}_k + \hat{\nu}$ is one of the upper bounds of $\boldsymbol{e}^\top \boldsymbol{y}_k$, and $\hat{d}_k$ is a downside deviation of $\boldsymbol{e}^\top \boldsymbol{y}_k$. Let

$$a := \sum_{j=1}^{n} \left| \boldsymbol{e}^\top \boldsymbol{y}_j - \left(\hat{\boldsymbol{v}}^\top \boldsymbol{x}_j + \hat{\nu}\right) \right|, \tag{1}$$

then we have

$$a = \sum_{j=1}^{n} \hat{d}_j.$$

Theorem 1. *Let $(\hat{\boldsymbol{d}}, \hat{\boldsymbol{v}}, \hat{\nu})$ be an optimal solution to the DLS problem. The upper bound $\hat{\boldsymbol{v}}^\top \boldsymbol{x}_j + \hat{\nu}$ of $\boldsymbol{e}^\top \boldsymbol{y}_j$ is independent of the choice of optimal solutions to the DLS problem. The same is true of the equation (1).*

3. A multiplier DEA model with downside-deviation restrictions

We define

$$V = \left\{ (\boldsymbol{v}, \nu) \;\middle|\; \begin{array}{l} \sum_{j=1}^{n} \left(\boldsymbol{v}^\top \boldsymbol{x}_j + \nu - \boldsymbol{e}^\top \boldsymbol{y}_j \right) \leq a \\ \boldsymbol{e}^\top \boldsymbol{y}_j \leq \boldsymbol{v}^\top \boldsymbol{x}_j + \nu, \ \forall j \\ \boldsymbol{v} \geq \boldsymbol{0} \end{array} \right\}$$

as a set of input multiplier $(\boldsymbol{v}, \nu)$. Here, $\boldsymbol{v}^\top \boldsymbol{x}_k + \nu$ is referred to as a substantial medal total of DMU_k for some $(\boldsymbol{v}, \nu) \in V$.

To evaluate the efficiency of Olympic achievements based on the substantial medal totals, we propose the following DEA model (LP_θ) :

$$\begin{array}{lll} \theta_k = \max. & \dfrac{\boldsymbol{u}^\top \boldsymbol{y}_k}{\boldsymbol{v}^\top \boldsymbol{x}_k + \nu} \\ \text{s.t.} & \boldsymbol{u}^\top \boldsymbol{y}_j - \left(\boldsymbol{v}^\top \boldsymbol{x}_j + \nu \right) \leq 0, \ \forall j \\ & \boldsymbol{e}^\top \boldsymbol{y}_j - \left(\boldsymbol{v}^\top \boldsymbol{x}_j + \nu \right) \leq 0, \ \forall j \\ & \displaystyle\sum_{j=1}^{n} \left(\boldsymbol{v}^\top \boldsymbol{x}_j + \nu - \boldsymbol{e}^\top \boldsymbol{y}_j \right) \leq a \\ & R^\top \boldsymbol{u} \geq \boldsymbol{0}, \boldsymbol{u} \geq \boldsymbol{0}, \boldsymbol{v} \geq \boldsymbol{0}, \end{array}$$

where $R = \begin{pmatrix} 1 & 0 & 0 \\ -P & 1 & 0 \\ 0 & -Q & 1 \end{pmatrix}$. The notations P and Q represent the lower bounds of the ratios u_1/u_2 and u_2/u_3, respectively. If $\theta_k = 1$, DMU_k is efficient. Otherwise, DMU_k is inefficient.

The efficiency of Olympic achievements for DMU_k in model (LP_θ) can be decomposed as

$$\frac{\boldsymbol{u}^\top \boldsymbol{y}_k}{\boldsymbol{v}^\top \boldsymbol{x}_k + \nu} = \frac{\boldsymbol{e}^\top \boldsymbol{y}_k}{\boldsymbol{v}^\top \boldsymbol{x}_k + \nu} \cdot \frac{\boldsymbol{u}^\top \boldsymbol{y}_k}{\boldsymbol{e}^\top \boldsymbol{y}_k}.$$

Here, the first component, $\boldsymbol{e}^\top \boldsymbol{y}_k / \left(\boldsymbol{v}^\top \boldsymbol{x}_k + \nu \right)$ is referred to as *the achievement ratio of substantial medal total*. The second component $\boldsymbol{u}^\top \boldsymbol{y}_k / \boldsymbol{e}^\top \boldsymbol{y}_k$ is termed as the *unit value index of medals*.

4. Examining the target of 30 gold medals in Tokyo 2020

The input vector of Japan $(\boldsymbol{x}_J)$ in Rio 2016 is used for predicting the total number of gold medals in Tokyo 2020, which is $\boldsymbol{x}_J = (x_{J1}, x_{J2}, x_{J3})^\top = (32477.2, 126958472, 41)^\top$. For the input vector $\boldsymbol{x}_J$, we derive the output possibility set from the dual problem of (LP_θ), which is

$$\mathcal{P}(\boldsymbol{x}_J) = \left\{ \boldsymbol{y} \;\middle|\; \begin{array}{l} \boldsymbol{y} \leq \sum_{j=1}^{n} \lambda_j \boldsymbol{y}_j - R\boldsymbol{d} \\ \boldsymbol{x}_J \geq \sum_{j=1}^{n} (\lambda_j + \mu_j) \boldsymbol{x}_j \\ \sum_{j=1}^{n} (\lambda_j + \mu_j) = 1 \\ -a\mu_j \leq \sum_{h=1}^{n} \mu_h (\boldsymbol{e}^\top \boldsymbol{y}_h) \\ \boldsymbol{d} \geq \boldsymbol{0}, \ \boldsymbol{\lambda} \geq \boldsymbol{0}, \ \boldsymbol{\mu}\text{:free} \end{array} \right\}.$$

Provided that the target of 30 gold medals in Tokyo 2020 is feasible, the optimal value of

$$\max \{ y_1 \mid \boldsymbol{y} \in \mathcal{P}(\boldsymbol{x}_J), \ \boldsymbol{y} \text{ is nonnegative} \} \qquad (2)$$

is equal or more than 30. However, the optimal value of the problem (2) is 26, which implies the target of 30 gold medals in Tokyo 2020 is infeasible even if we investigate all possible combinations of medals within $\mathcal{P}(\boldsymbol{x}_J)$.

How many medals will Japan win if 26 gold medals are actually achieved? To answer this question from a practical point of view, it is necessary characterize the possible patterns of medal won based on the patterns of past Olympic games. Consider the following linear combination:

$$\boldsymbol{Y}_J = \left\{ \sum_{t=1}^{4} \gamma^t \boldsymbol{y}_J^t \;\middle|\; \gamma^1 \geq 0, \cdots, \gamma^4 \geq 0 \right\},$$

where $\gamma^1, ..., \gamma^4$ are scalars and $\boldsymbol{y}_J^1, \ldots, \boldsymbol{y}_J^4$ are integer vectors of medals that Japan won in the past four Olympic games. Assume the medals Japan will achieve in Tokyo 2020 belong to $\boldsymbol{Y}_J$. Therefore, if Japan bags 26 gold medals in Tokyo 2020, the maximal number of medal totals will be the optimal solution of the following problem:

$$\max \left\{ \sum_{r=1}^{3} y_r \;\middle|\; \boldsymbol{y} \in \mathcal{P}(\boldsymbol{x}_J), \ \boldsymbol{y} \text{ is nonnegative} \right\}. \qquad (3)$$

We found that the optimal value of the problem (3) is a total of 63, and $\boldsymbol{y}^* = (y_1^*, y_2^*, y_3^*)^\top = (26, 15, 33)^\top$. Thus, Japan will win 63 medals at most, which consists of 26 gold, 15 silver, and 22 bronze medals.