Economics Bulletin

Volume 35, Issue 2

A medal share model for Olympic performance

Ang Sun
Renmin University of China

Rui Wang
Tsinghua University

Zhaoguo Zhan
Tsinghua University

Abstract

A sizable empirical literature relates a nation's Olympic performance to socioeconomic factors by adopting linear regression or a Tobit approach suggested by Bernard and Busse (2004). We propose an alternative model where a nation's medal share depends on its competitiveness relative to other nations and the model is logically consistent. Empirical evidence shows that our model fits data better than the existing linear regression and Tobit model. Besides Olympic Games, the proposed model and its estimation method could also be similarly applied to other settings with competitions.

Citation: Ang Sun and Rui Wang and Zhaoguo Zhan, (2015) "A medal share model for Olympic performance", *Economics Bulletin*, Volume 35, Issue 2, pages 1065-1070
Contact: Ang Sun - ang.sun@gmail.com, Rui Wang - wangr.11@sem.tsinghua.edu.cn, Zhaoguo Zhan - zhanzhg@sem.tsinghua.edu.cn.
Submitted: June 04, 2014. **Published:** April 22, 2015.

1 Introduction

There exists a sizable literature that relates Olympic performance of participating nations to their socioeconomic factors. The leading analytical approaches in this literature include the commonly used linear regression model and the Tobit model. For example, Hoffmann et al. (2002) and Johnson and Ali (2004), among many others, use the number of medals as the proxy for Olympic performance, and relate it to socioeconomic factors by adopting the linear regression model, which is then estimated by Ordinary Least Squares (OLS). On the other hand, Bernard and Busse (2004) use the share of medals as the dependent variable, and apply the Tobit method to account for the fact that medal shares are censored at zero.

However, there are three features of Olympic Games that are at least partially ignored by the existing approaches. First, the number of medals a nation can win depends on its own competitiveness as well as the competitiveness of other nations. Second, medal shares in each Olympiad lie between 0 and 1 and sum up to unity. Third, competition for medals is mostly among more competitive nations in Olympic Games, rather than all participating nations. The linear regression model commonly ignores all the features above. By contrast, Bernard and Busse (2004) take relative competitiveness into consideration in their Tobit model for Olympic medal shares. Nevertheless, there is no guarantee that medal shares in the Tobit model lie between 0 and 1 and sum up to unity. Furthermore, Bernard and Busse (2004) do not distinguish between nations with medals and nations without medals in Olympic competition.

In this paper, we propose a simple model to incorporate all the three features above. As in Bernard and Busse (2004), we also focus on the medal share instead of the medal count. The subtle difference between our model and Bernard and Busse (2004) is that we incorporate the feature that the competitions for medals are generally among competitive nations. We view the collection of Olympic medals as a market that nations compete to occupy, and each competitive nation ends up with some share of the market of medals, depending on their relative competitiveness. Although nations without any medal do participate in Olympic Games, their role in determining medal shares of competitive nations is negligible. Thus, we divide nations into two groups at the first stage, depending on whether they are likely to win at least one medal. The medal share in our model is guaranteed to lie between 0 and 1, and sums up to unity. At the second stage, we model the medal shares following the market share theorem in Bell et al. (1975) and the market attraction models in, e.g, Ghosh et al. (1984) and Fok et al. (2002). Empirical results show that our model fits data well, since the predicted numbers of medals based on the proposed model are closer to the actual observations, compared to the predictions based on the linear regression and the Tobit model.

Although our model is proposed in the Olympic setup, it could also be applied to other similar settings. For example, just as nations compete for a large medal share in Olympiads, companies also compete for market shares in industry. Thus in principle, our model can be similarly applied to the marketing literature.

2 Model and Estimation

2.1 Model

Consider in the t^{th} $(t = 1, ..., T)$ Olympiad, there are N_t participating nations. Let m_{it} be the number of medals won by nation i at time t, then the corresponding medal share, denoted by ms_{it}, is given by $ms_{it} = m_{it}/\sum_{j=1}^{N_t} m_{jt}$.

Similar to Trivedi and Zimmer (2013), we employ a Probit model to describe whether nation i is actively involved in medal competition (i.e., wins at least one medal) in the Olympiad at time t:

$$Y_{it} = \mathbf{1}(\mathbf{Z}_{it}'\boldsymbol{\gamma} + u_{it} > 0) \tag{1}$$

where $u_{it} \sim NID(0, 1)$, $\mathbf{Z}_{it}$ is the column vector of variables (see Trivedi and Zimmer (2013) for the discussion) that help explain the outcome Y_{it}, and $\boldsymbol{\gamma}$ is the associated vector of parameters. $Y_{it} = 0$ means that nation i wins no medal at time t.

Equation (1) thus divides participating nations into two groups for each Olympiad. One group with $Y_{it} = 0$ does not effectively affect Olympic medal share at time t, since $ms_{it} = 0$ in this group; by contrast, the other group with $Y_{it} = 1$ and $ms_{it} > 0$ effectively determines how medals are shared.

Take the 2012 London Summer Olympics for example. In total, 204 nations participated in this Olympiad, while only 85 nations won at least one medal. The large number of nations without any medal has motivated our model, which will impose the negligible role of these nations in affecting medals shares of more competitive nations, as shown below.

For nations that can win at least one medal, we assume their medal shares depend on the relative competitiveness (see, e.g., Bell et al. 1975):

$$ms_{it} = \frac{C_{it}}{\sum_{j=1}^{n_t} C_{jt}}, \;\; i = 1, ..., n_t < N_t, \;\; t = 1, ..., T. \tag{2}$$

where C_{it} is the competitiveness of nation i at time t, n_t is the number of nations that win at least one medal at time t. For logical consistency (see, e.g., Bultez and Naert 1975), i.e., $0 < ms_{it} < 1$, and $\sum_{i=1}^{n_t} ms_{it} = 1$, we need to impose $C_{it} > 0$.

For analytical purposes, we describe C_{it} by a function $C_{it} = f(\mathbf{X}_{it}, v_{it})$, where $\mathbf{X}_{it}$ is the $k \times 1$ vector of factors that determine competitiveness of nation i at time t, and v_{it} is the random shock in Olympic Games (e.g., referee error). We assume v_{it} has mean zero and is independently distributed across i and t; in addition, v_{it} is allowed to be correlated with the first step error u_{it} in Equation (1). Furthermore, since there is no theoretical guidance on the form of f, we adopt the following simplification:

$$C_{it} = f(\mathbf{X}_{it}, v_{it}) = e^{\beta_0 + \mathbf{X}_{it}'\boldsymbol{\beta} + v_{it}} \tag{3}$$

The motivation of (3) is as follows. First, it is common practice to combine factors of $\mathbf{X}_{it}$ in the linear manner for convenience. Second, the exponential transformation of the linear combination of factors is to make $C_{it} > 0$, which also makes model estimation easy to implement, as we show in the next section.

Plugging the expression of C_{it} in (3) into (2), and combining with Equation (1), we get the following medal share model for Olympic Games:

$$ms_{it} = \begin{cases} \frac{e^{\mathbf{X}'_{it}\boldsymbol{\beta}+v_{it}}}{\sum_{j=1}^{n_t} e^{\mathbf{X}'_{jt}\boldsymbol{\beta}+v_{jt}}}, & Y_{it} = 1; \\ 0, & Y_{it} = 0. \end{cases} \tag{4}$$

where $Y_{it} = \mathbf{1}(\mathbf{Z}'_{it}\boldsymbol{\gamma} + u_{it} > 0)$ results from the Probit model in (1).

2.2 Estimation

Given the model of (4) above, the interest is to conduct estimation with the data for $ms_{it}, \mathbf{X}_{it}, \mathbf{Z}_{it}$. To simplify the estimation, we linearize the model in the manner of Nakanishi (1972), i.e., take the natural log of (4) when $Y_{it} = 1$, and then remove the sample average, we get:

$$\ln(\frac{ms_{it}}{\overline{ms_t}}) = \begin{cases} (\mathbf{X}_{it} - \overline{\mathbf{X}_{it}})'\boldsymbol{\beta} + v_{it} - \overline{v_t}, & Y_{it} = 1; \\ missing, & Y_{it} = 0. \end{cases} \tag{5}$$

where $\overline{ms_t} = (ms_{1t}ms_{2t}...ms_{n_tt})^{1/n_t}$, $\overline{\mathbf{X}_{it}} = \frac{1}{n_t}\sum_{i=1}^{n_t}\mathbf{X}_{it}$, $\overline{v_t} = \frac{1}{n_t}\sum_{i=1}^{n_t} v_{it}$.

It is worth emphasizing that we now encounter a missing data problem, i.e., when $Y_{it} = 0$, $\ln(\frac{ms_{it}}{\overline{ms_t}})$ does not exist. So the dependent variable in (5) is only available when $Y_{it} = 1$.

To put it in a different manner, the simplified model in (5) is close to an OLS model except that $\ln(\frac{ms_{it}}{\overline{ms_t}})$ exists only when $Y_{it} = 1$ (or equivalently, $ms_{it} > 0$); by contrast, $\ln(\frac{ms_{it}}{\overline{ms_t}})$ does not exist, when $Y_{it} = 0$ (or equivalently, $ms_{it} = 0$). As a result, we encounter a setup similar to the one in Heckman (1979). The correction proposed by Heckman (1979) provides us an appropriate solution to address the issue of missing values in (5).[1]

3 Application

3.1 Data

In the empirical application, we use the data of seven Summer Olympiads from 1988 to 2012 to measure the Olympic performance of participating nations. The data of medal shares (ms_{it}) is from the official web site of the International Olympic Committee.

For socioeconomic factors of $\mathbf{X}_{it}$, we follow the literature to consider population, per capita GDP, per capita health expenditure, climate and a dummy for the political system. Specifically, we use the data of population, per capita GDP and per capita health expenditure (in natural logs) from the web site of World Bank. Climate is the percentage of areas in a nation where frost days exceed 20 days in a year, and the data source is as in Masters and McMillan (2001). Since the Summer Olympic Games typically take place around the middle of a year, we use the socioeconomic data one year before the Olympic

[1]Different from Heckman (1979), our missing data problem is due to log-linearization, while medal shares are themselves observable. By contrast, the missing data problem in Heckman (1979) is due to unobservable data. Thus our model setup does not coincide Heckman (1979), although Heckman (1979)'s method is suitable for our estimation.

year. We use a dummy variable to distinguish the political system: it equals 1 for communist nations and 0 otherwise. In addition to these socioeconomic factors, we also employ a dummy variable for winning no medal at all from last Olympiad at the first stage for $\mathbf{Z}_{it}$, in order to deal with Olympic performance triggered by ethnic characteristics or nation specific factors.

Missing observations are mainly due to the data of health expenditure and climate. Our estimation sample includes around 200 nations for each Olympiad and a total of 1418 observations.

3.2 Estimation Outcome

In Table 1, we explicitly compare the outcome of the three approaches described in this paper.

Table 1: Relating Olympic Performance to Socioeconomic Factors

	I. Medal Count (OLS)	II. Medal Share (Tobit)	III. log of Medal Share
Model Estimation			
Per Capita GDP	2.688***	0.006***	0.300***
	(0.262)	(0.000)	(0.040)
Population	3.412***	0.009***	0.430***
	(0.369)	(0.000)	(0.034)
Communism	16.128***	0.020***	1.128***
	(5.090)	(0.004)	(0.263)
Health Expenditure	4.575***	0.008***	0.345**
	(1.234)	(0.002)	(0.150)
Climate	5.957***	0.018***	0.796***
	(1.106)	(0.002)	(0.115)
Intercept	−58.877***	−0.179***	-
	(6.347)	(0.012)	
Model Comparison			
R-squared	0.395	-	0.517
Squared Correlation	0.395	0.618	0.670
Mean Absolute Error	6.846	3.933	3.477

Note: ***, **, * stand for significance at 1%, 5% and 10%. Standard errors are in brackets.

The first approach is to relate medal counts to socioeconomic factors by OLS (see, e.g., Hoffmann et al. 2002), while the second approach uses medal shares as the dependent variable, and Tobit censored at zero as the estimation method (see, e.g., Bernard and Busse 2004). Both of these two approaches are commonly seen in the existing studies on Olympic performance. The third approach is the one proposed in this paper, where the dependent variable is the natural logarithm of medal shares.

Table 1 presents the outcome of the three approaches one by one. For Column I, we use the medal count (denoted by m_{it} above) as the dependent variable for OLS. For Column II, we use the medal share (denoted by ms_{it} above) as the dependent variable, and conduct the Tobit estimation. For Column III, we use the natural logarithm of the medal share (denoted by $\ln(\frac{ms_{it}}{ms_t})$ above, after removing the sample average) as the dependent variable, and conduct Heckman (1979)'s estimation.

As shown in Table 1, all three approaches suggest that the listed socioeconomic factors are crucial for Olympic success. Note that since the dependent variables differ, the magnitudes of the reported coefficients in Table 1 are not directly comparable, but the signs for the coefficients of socioeconomic factors all remain positive.

3.3 Model Comparison

To compare the empirical performance of the three approaches (models) presented in Table 1, we calculate the predicted medal count, denoted by $\hat{m}_{it}$, based on each approach. Given the predicted medal count $\hat{m}_{it}$ and the observed medal count m_{it} in Olympic Games, we compute their squared correlation for each approach:

$$Squared\ Correlation = \rho(\hat{m}_{it}, m_{it})^2 \tag{6}$$

In addition, as a check of sensitivity, we report the mean absolute error of each approach, representing the absolute difference between the predicted and the actual medal counts on average:

$$Mean\ Absolute\ Error = \frac{1}{\sum_{t=1}^{T} N_t} \sum_{t=1}^{T} \sum_{i=1}^{N_t} |\hat{m}_{it} - m_{it}| \tag{7}$$

The bottom panel of Table 1 presents the squared correlation and mean absolute error for the three approaches. The largest squared correlation and least mean absolute error associated with our model indicate that the proposed model best fits data.

We also conduct the out-of-sample prediction by removing the latest two Olympiads and use the resulting estimands to forecast for the two most recent Olympic outcomes. We find that in the prediction, our medal share model still performs the best with the highest squared correlation and the least mean absolute error, compared to OLS and Tobit.[2]

4 Conclusion

We propose a medal share model to incorporate the features of competition in Olympic Games, which are commonly ignored by existing studies. Empirical evidence shows that the proposed model similarly favors the leading socioeconomic factors but fits data better than the popular linear regression model and Tobit model in the literature.

[2] For out-of-sample, the Squared Correlation of our model is 0.707, compared with 0.604 of the Tobit and 0.398 of the OLS. The MAE of our model is 2.050, compared with 5.201 and 7.968 when using the Tobit and OLS, respectively.

References

Bell, D. E., Keeney, R. L., and Little, J. D. 1975. A market share theorem. *Journal of Marketing Research*, pages 136–141.

Bernard, A. B. and Busse, M. R. 2004. Who wins the Olympic Games: Economic resources and medal totals. *Review of Economics and Statistics*, 86(1):413–417.

Bultez, A. V. and Naert, P. A. 1975. Consistent sum-constrained models. *Journal of the American Statistical Association*, 70(351a):529–535.

Fok, D., Franses, P. H., and Paap, R. 2002. *Econometric analysis of the market share attraction model*, volume 16. Emerald Group Publishing Limited.

Ghosh, A., Neslin, S., and Shoemaker, R. 1984. A comparison of market share models and estimation procedures. *Journal of Marketing Research*, pages 202–210.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica*, pages 153–161.

Hoffmann, R., Ging, L. C., and Ramasamy, B. 2002. Public policy and Olympic success. *Applied Economics Letters*, 9(8):545–548.

Johnson, D. K. and Ali, A. 2004. A Tale of Two Seasons: Participation and Medal Counts at the Summer and Winter Olympic Games. *Social Science Quarterly*, 85(4):974–993.

Masters, W. A. and McMillan, M. S. 2001. Climate and scale in economic growth. *Journal of Economic Growth*, 6(3):167–186.

Nakanishi, M. 1972. Measurement of sales promotion effect at the retail level-a new approach. In *Proceedings, Spring and Fall Conferences, American Marketing Association*, volume 338, pages 303–311.

Trivedi, P. K. and Zimmer, D. M. 2013. Success at the Summer Olympics: How Much Do Economic Factors Explain? Working paper.