

Introduction

Despite the predictive accuracy of the previous model, documented [here](#), we found some room for improvement by a) updating the data and b) re-estimating the model to find new important characteristics.

Predictions, along with some explanatory information, are presented in the press release dated March 12, 2012, hosted [here](#).

Data

All data used for this analysis are available [here](#). We now have data on the following characteristics of every possible participant nation, for every Olympic year between 1952 and 2012. There are some clear and unfortunate omissions (e.g. Belarus, Ukraine, and many republics of the former USSR are lacking the data that we would need to estimate medal counts). Therefore, we are only able to predict medal counts for 130 nations in 2012. The econometric estimation of our model shows a 96% correlation with actual Summer Games medal counts over that entire period, 95% for gold medals alone.

In contrast with the previously published version of this model, we elected to omit two national characteristics: political structure and climate. Political structure has clearly had an important role historically, as our previous model showed that Communist and single-party regimes have enjoyed 18 more medals than their democratic peers (7 of them gold). However, in the current international environment there are simply too few remaining examples to permit any inference for predictive purposes. Climate showed a small but significant effect in the past, but in this new calibration its effects were negligible at best.

Instead, we have incorporated two new effects which appear to explain (and perhaps predict) medal counts: a host nation effect that both pre-dates and post-dates the actual hosting of the Games, and a “nation-specific cultural effect”. Furthermore, we permitted the importance of every national attribute to change over time, choosing 1992 as the point of re-estimation (a year of dramatic reconfiguration of the Olympic map, in the aftermath of the Soviet Union). In this way, we can investigate whether certain characteristics were more important pre-1992 or in recent years.

We define the pre-hosting effect as applying to a nation in the venue that precedes its own hosting, 4 years in advance. The assumption is that since a nation knows that it will host a full 6 years in advance, it may choose to prepare infrastructure or athletes well in advance of its actual hosting duties.

There may also be a legacy effect, or post-hosting effect, as a former host enjoys the benefits of the infrastructure and athletes (now potentially world champions or coaches or advisors) for venues that follow the hosting duties. We estimate this effect for two venues (8 years) after the host’s physical obligations.

We also estimate a cultural effect, specific to each nation, permitting the data to determine whether immeasurable effects are present. This effect will hopefully capture some of the differences between nations which we find impossible to measure or quantify: financial support for athletics, a culture of sport and/or competition, prevalence of doping, legacies of previous athletic excellence, and other effects.

Income per capita data are corrected for purchasing power parity, and expressed in thousands of real 2005 US dollars using the Penn World Tables (PWT). Income is measured in the year preceding the Games (for logistical reasons, as we simply cannot know the income any nation by February of that year, when the Winter Games are traditionally hosted, nor even by July/August when the Summer Games are traditionally held). To extrapolate income to 2011, we calculated percentage changes in real income using the Stateman's Yearbook, and applied those changes to the (most recent available) 2009 values of the PWT.

Population comes from the PWT as well, augmented by the Stateman's Yearbook for 2011, and is expressed in billions.

Model

The model now takes four formats, as shown here. All specific values are the result of econometric estimation using all 1952-2010 data to fit existing data as closely as possible.

a) for all medals, pre-1992

$$\begin{aligned} \text{medals} = & 1.48 + [1.15 \cdot 10^{-3} \cdot \text{total medals available}] \\ & + [0.09 \cdot \text{income} + 3.33 \cdot 10^{-3} \cdot \text{income}^2] \\ & + [-4.46 \cdot \text{population} + 32.66 \cdot \text{population}^2] \\ & + [22.05 \text{ if home nation currently} + 1.56 \text{ if home nation in next Games} + \\ & \quad + 9.53 \text{ if home nation in either of previous two Games}] \\ & + [2.59 \text{ if physical neighbor to current hosting nation}] \\ & + \text{estimated nation-specific effect which varies by nation} \end{aligned}$$

b) for all medals, 1992-2012

$$\begin{aligned} \text{medals} = & 2.23 + [-6.11 \cdot 10^{-4} \cdot \text{total medals available}] \\ & + [0.07 \cdot \text{income} - 2.27 \cdot 10^{-4} \cdot \text{income}^2] \\ & + [86.42 \cdot \text{population} - 24.28 \cdot \text{population}^2] \\ & + [18.06 \text{ if home nation currently} + 9.33 \text{ if home nation in next Games} + \\ & \quad + 4.70 \text{ if home nation in either of previous two Games}] \\ & + [1.24 \text{ if physical neighbor to current hosting nation}] \\ & + \text{estimated nation-specific effect which varies by nation} \end{aligned}$$

c) for gold medals, pre-1992

$$\begin{aligned} \text{medals} = & 0.42 + [6.23 \cdot 10^{-4} \cdot \text{total gold medals available}] \\ & + [0.09 \cdot \text{income} + 7.88 \cdot 10^{-4} \cdot \text{income}^2] \\ & + [-4.24 \cdot \text{population} + 11.60 \cdot \text{population}^2] \\ & + [10.37 \text{ if home nation currently} - 0.38 \text{ if home nation in next Games} + \\ & \quad + 2.78 \text{ if home nation in either of previous two Games}] \\ & + [0.41 \text{ if physical neighbor to current hosting nation}] \\ & + \text{estimated nation-specific effect which varies by nation} \end{aligned}$$

d) for all medals, 1992-2012

$$\begin{aligned} \text{medals} = & 0.33 + [2.71 \cdot 10^{-3} \cdot \text{total gold medals available}] \\ & + [0.02 \cdot \text{income} - 5.49 \cdot 10^{-5} \cdot \text{income}^2] \\ & + [0.24 \cdot \text{population} + 19.02 \cdot \text{population}^2] \\ & + [11.91 \text{ if home nation currently} + 3.85 \text{ if home nation in next Games} + \\ & \quad + 3.35 \text{ if home nation in either of previous two Games}] \\ & + [0.29 \text{ if physical neighbor to current hosting nation}] \\ & + \text{estimated nation-specific effect which varies by nation} \end{aligned}$$

The single most potent factor appears to be the host nation advantage, which confers additional 37 medals (22 of them gold) on the host, roughly half of which occur in the hosting year. That advantage has increased by 20% in the years since 1992.

Income has become less important, and population has become more important, since 1992. This is unsurprising given the rising importance of China as an Olympic power.