"Prediction is very difficult, especially if it's about the future."
– Nils Bohr, Nobel laureate in Physics.

The Olympic medals ranking: Does the past predict the future?

Julia Bredtmann, Carsten J. Crede, Sebastian Otten

When the 31st Olympic Games start on the 5th of August this year in Rio de Janeiro, the world will be watching athletes from more than 200 nations competing for medals. The Summer Olympics are the most prestigious sports event in the world and an estimated 3.6 billion people have watched the last Olympic Summer Games.[1] If you followed the past few Games as well, you might have noticed that most medals were won by only a few countries including the United States, China, and Russia. Indeed, a closer look at the nation rankings for Olympic medals reveals that these same nations tend to win the most medals at all Olympic Games.

Thus, one might be tempted to ask by how much winning a medal is determined by the individual athlete's ability and talent and how much is driven by the conditions in his or her country of origin? A large number of academic studies have attempted to address this question and explain the phenomenon of a nation's persistent Olympic success.

The determinants of success

These studies have found that a number of socio-economic variables are reliable predictors of Olympic success. For example, a country's gross domestic product (GDP) and its population size are strongly correlated with Olympic success.[2] Of course, the GDP by itself has no impact on an athlete's performance. However, it correlates with a large number of factors that affect an athlete's ability to train, and as such is a so-called proxy variable for these other factors: in a wealthy country, the population can dedicate more time to leisure activities such as sports, is healthier, and can afford professional athletes as well as a better sports infrastructure for more effective training. Another important variable is a country's population size; provided that world-class athletic talent is uniformly distributed across the world's population, larger countries should generally produce more top athletes.

Figures 1 and 2 show the correlation between a country's GDP per capita and population size, respectively, and the total number of medals won at the 2012 Olympic Games for the 30 top-scoring countries. As can be seen from the linear regression line, larger and wealthier countries generally win more medals. However, especially Figure 1 shows the existence of some outliers. For example, the USA, China, and Russia have won more medals at the 2012 Games than one would expect from their GDP per capita.

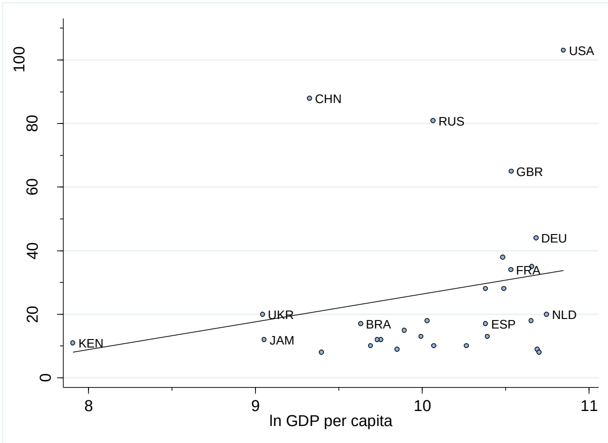

Figure 1: Correlation between GDP per capita and total medals

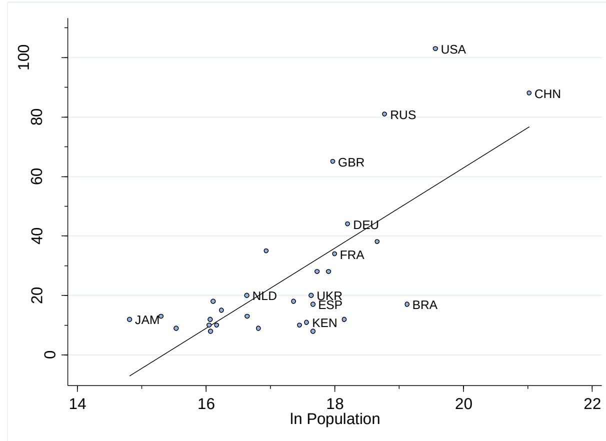

Figure 2: Correlation between population and total medals

Other predictors include a country's present or past political system (autocratic nations and planned economies tend to make greater investments in athletes to obtain prestige, and countries formerly featuring such systems continue to profit from their established systems to promote sports), whether it is the (next) host country of the Games (host and upcoming host nations invest more in athletes before the Games and send more athletes, and the host country's athletes benefit from the home advantage, e.g. they are accustomed to the climatic conditions), and whether women are equally likely to participate in sports and be able to train to become athletes.[3,4]

Predicting Olympic medals

Exploiting the correlation between socio-economic variables and Summer Olympic success as well as the persistence in Olympic success over time, it is possible to reliably predict the Summer Olympic medals ranking based on the overall number of medals won for the upcoming games. This can be done using

Out-of-sample predictions

Consider a simple regression model,

$$Y_{it} = \beta_1 + \beta_2 X_{it} + \varepsilon_{it},$$

where Y is the outcome variable, X is the explanatory variable, ε is the error term, and indices *i* and *t* denote countries and time, respectively. Assume that we have three time periods: Two in the past and one in the future. To predict Y_{it} in the future period, we can estimate the above model using only observations from the two previous periods, i.e. t = [1, 2], and obtain the corresponding estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$. Using the estimated past relationship between X and Y in periods 1 and 2 and assuming that this relationship holds for the future period, we can predict the values of the outcome variable in period 3 by plugging in values of X in period 3 into

$$\hat{Y}_{i3} = \hat{\beta}_1 + \hat{\beta}_2 X_{i3},$$

to obtain predictions of the outcome variable for each country *i*.

regression analysis to conduct out-of-sample predictions (see box on page XX).

Before carrying out medal predictions for the 2016 Olympics in Brazil, we need to develop a model that will perform well in predicting Olympic medals. For this purpose, we predict medals for the 2012 Olympics with two different models established using pre-2012 Olympic medal counts and then compare these predictions with the actual medals won by nations. To do this, we first fit the regression models based on the results at the Olympic Games between 1996 and 2008, i.e. using 676 country-year observations, and then carry out the out-of-sample predictions for 2012. We do not consider Olympic Games before 1992 (neither as current nor as lagged observations) to avoid the huge structural change caused by the fundamental social and political changes taking place around 1990 to confound our results.

The first model we consider, which we call the *naive* model, primarily describes a country's medals as a function of the number of medals won at the previous Olympic Games, i.e. a nation's medals at the 2008 Games are predicted primarily based on how many medals it won at the Games in 2004. This captures the persistence in performance of countries. The other explanatory variable in this model is a linear time trend capturing the years of the different Olympic Games; this is done to take into account that the

The estimated models

Both models are estimated using Ordinary Least Squares (OLS). The naive model is:

$$Medals_{it} = \beta_1 + \beta_2 Lag\ Medals_{it} + \beta_3 Year_t + \varepsilon_{it},$$

where *Medals* denotes the total number of medals (incl. gold, silver, and bronze medals) won by a country at the corresponding Olympic Games, *Lag Medals* contains a country's medals at the preceding Games, and *Year* denotes the year of the Olympic Games. The latter variable is included to capture the steady increase in the total number of medals awarded at the Olympic Games over time.

The sophisticated model includes additional explanatory variables that capture country-specific characteristics:

$$Medals_{it} = \gamma_1 + \gamma_2 Lag\ Medals_{it} + \gamma_3 ln\ GDP_{it} + \gamma_4 \ln Pop_{it} + \gamma_5 Host_{it} + \gamma_6 Next\ Host_{it} + \gamma_7 Planned_{it} + \gamma_8 Muslim_{it} + \gamma_9 Year_t + \zeta_{it},$$

where *ln GDP* denotes the natural logarithm of a country's GDP per capita and *ln Pop* it is the natural logarithm of a country's population. Both variables are included in logs to acknowledge that the positive effects of GDP per capita and population size on Olympic medals diminish with increasing values of these variables. *Host* and *Next Host* are indicator variables for the current host and the upcoming host country. *Planned* is an indicator variable denoting whether a country has or had a fully centralised planned economy (such as former members of the Soviet Union, China, and Cuba) and controls for higher expenditure of such countries for sports to promote national prestige. In case of a country's switch to another economic system, it measures the extent to which it still profits from the past sports infrastructure. Finally, *Muslim* is an indicator for countries with a predominant (>50%) Muslim population, which tend to send less female athletes[4], and tend to have a lower share of the female population active in professional sports. For all explanatory variables considered, we do not assume them to have a causal impact on a country's Olympic success, but are aware that they might capture both direct effects and indirect effects of other, unobserved country-specific factors.

overall number of medals increases over time because of an increasing number of events over time.

The second model that we estimate, which we label the *sophisticated* model, expands the naive model with additional socio-economic variables. These include a country's GDP per capita and population, whether the country is or was a planned economy, whether the country is the host (for 2012 this was the United Kingdom) or upcoming host (in 2012 this was Brazil), and whether it has a predominant Muslim population (such countries tend to send less female athletes and win fewer medals in women's events). For more details, see the box on page XX.

To assess whether the sophisticated model outperforms the naive model, we can calculate measures of fit and compare them between the two models. For this purpose, we calculate the mean absolute error (MAE) and the mean forecast error (MFE) for both the naive and the sophisticated model (see box on page XX).

For the full sample including 181 countries, the MAE is 1.43 for the naive model, and 1.41 for the sophisticated model. In other words, on average, in the naive model predictions are off by 1.43 medals from the true number of medals won by each country, and by 1.41 in the sophisticated model. Considering the top 15 countries only, the difference in MAE is more pronounced with a reduction from 6.6 to 5.8 when switching from the naive to the sophisticated model. At the same time, the predictions in the sophisticated model are subject to a lower uncertainty, as the MFE is 3.2 in the sophisticated model compared to 3.4 in the naive model. Therefore, these results suggest that including socio-economic variables into the model slightly increases the precision of the predictions.

Mean absolute error (MAE) and mean forecast error (MFE)

The mean absolute error is a measure of the average inaccuracy associated with a set of model-produced estimates. It compares the predicted values $\hat{y}$ for the outcome variable with the true values y across all individual estimates i, and calculates a measure of the absolute value of differences:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|.$$

The lower therefore the MAE, the better the fit of the model.

A second measure of prediction inaccuracy is the forecast error. The forecast error is the standard error of the point prediction and thereby expresses the uncertainty in estimating the unknown value of y for an individual observation with known X values. A forecast error can be obtained for each country's predicted medals. For the purpose of model comparison, we calculate the mean value of these errors for the top 15 countries.

Table 1 compares the actual results for the top 15 countries at the 2012 Olympics with the predictions based on the two different models. For the naive and the sophisticated model, it shows both the number of predicted medals as well as the absolute difference between the predicted and the actual medals won by each country. For the sophisticated model, it further shows the predicted rank.

Table 1: Actual and predicted medals for the 2012 Olympic Games

Country	Actual results		Naive model		Sophisticated model		
	2012 Rank	Medals	Predicted medals for 2012	Diff.	Predicted medals for 2012	Diff.	Predicted 2012 Rank
United States	1	103	110	7	105	2	1
China	2	88	100	12	96	8	2
Russia	3	81	73	8	70	11	3
United Kingdom	4	65	47	18	60	5	4
Germany	5	44	41	3	39	5	6
Japan	6	38	25	13	24	14	12
Australia	7	35	46	11	44	9	5
France	8	34	41	7	39	5	6
Italy	9	28	27	1	26	2	10
South Korea	9	28	31	3	30	2	8
Netherlands	11	20	16	4	16	4	17
Ukraine	11	20	27	7	27	7	9
Canada	13	18	19	1	19	1	14
Hungary	13	18	10	8	10	8	21
Brazil	15	17	15	2	26	9	10
Spain	15	17	18	1	18	1	15
Sum		654	646	106	649	93	
MAE				6.6		5.8	
MFE			3.4		3.2		

The table shows that the naive model already predicts Olympic success fairly well. In other words, due to the persistence in Olympic success, one can determine how many medals a country wins quite reliably by merely looking at the results at the last Olympic Games. In fact, due to the nearly constant number of awarded medals at the 2008 and 2012 Olympic Games, the naive model delivers exactly the same results as an "ultra-naive model", which obtains the number of medals won at the 2012 Games by simply multiplying the total number of medals to be awarded in 2012 by the proportion of medals won by each nation in the 2008 Games. In total, the naive model predicts 106 medals of the 654 medals won by these countries wrong. The most notable outlier is the 2012 host United Kingdom, which won 18 more medals than predicted by the naive model that ignores the host country effect.

Turning to the results of the sophisticated model, it can be seen that it performs slightly better than the naive model. This is in line with the finding of smaller MAE and MFE values indicating a higher precision of the sophisticated model. The absolute difference between predicted and actual medals reduces from 106 to 93 medals. The increase in precision of predictions primarily arises from smaller differences for the 2012 host United Kingdom and the United States. For the United Kingdom, the difference goes down from a staggering 18 to only 5 medals. This is a result of controlling for the host country effect outlined above. Yet, for Brazil the prediction error increases from 2 to 9 medals. This follows from the fact that compared to other previous upcoming host nations, Brazil performed surprisingly worse at the 2012 Games. With its increase in medals won by only 2, Brazil stays 7 medals behind the expected advantage of being the upcoming host nation.

The 2016 predictions

Given that the sophisticated model performs better in predicting the 2012 Olympics, we use it to predict the Olympic medals for the top 15 nations at the 2016 Games.[5] To estimate the parameters of our model, we use the results at the Olympic Games between 1996 and 2012. The predictions can be found in Table 2.

The model predicts that the United States, China, Russia, and the United Kingdom will retain their top positions at the medals ranking.[6] Big winners for the 2016 Games are Brazil and Japan. As the host country, Brazil will send more athletes (431 Brazilian athletes are nominated to compete in 27 sports in 2016 compared to 258 in 24 sports in 2012) than it did to previous Games,[7,8] and in preparation of hosting will have invested more in the development of its national talent pool. Furthermore, Brazilian athletes will likely be the stadium visitors' favourites, receiving the most cheers and support from the audience. Japan will host the 2020 Summer Olympics. Therefore, the investment it makes in its athletes in preparation for the 2020 Games are expected to pay off already. The success of Brazil and Japan comes at the cost of most of the other countries in the top 15 table. Nevertheless, despite being predicted to lose 5 medals compared to 2012, the United States is predicted to retain its top position at the Olympics medals ranking.

Table 2: Predicted medals for the 2016 Olympic Games

Country	Predicted rank for 2016	Predicted medals for 2016	Rank in 2012	Medals in 2012	Diff. in ranks	Diff. in medals to 2012
United States	1	98	1	103	0	-5
China	2	84	2	88	0	-4
Russia	3	77	3	81	0	-4
United Kingdom	4	62	4	65	0	-3
Japan	5	46	6	38	1	8
Germany	6	42	5	44	-1	-2
Australia	7	33	7	35	0	-2
Brazil	7	33	15	17	8	16
France	7	33	8	34	1	-1
Italy	10	27	9	28	-1	-1
South Korea	10	27	9	28	-1	-1
Ukraine	12	20	11	20	-1	0
Netherlands	13	19	11	20	-2	-1
Canada	14	18	13	18	-1	0
Hungary	14	18	13	18	-1	0
MFE		3.1				

How much should we trust these predictions? There are only indirect ways to assess the quality of the predictions. One possibility is to look at the MFE for the 2016 predictions, which goes down to 3.1 for the 2016 Games from 3.2 for the 2012 Games. Therefore, the model predictions feature a similar level of uncertainty as those for the 2012 Olympic Games. Thus, it is likely that the accuracy of the 2016 Olympic medals prediction will be in the same range as those for the 2012 Games.

As we have seen, the suspicion that the same countries tend to win the majority of medals is well founded. As a result, even simple methods produce fairly accurate forecasts. More sophisticated models that incorporate socio-economic factors offer a way to improve these predictions. Yet, apart from the host and the upcoming host effects their added value is rather limited, as the effects of these factors are already reflected in the medals won at the previous Games. Thus, in the context of the Olympic Games, past success is indeed a good predictor of future success.

However, before you give your hopes up if you support "underdog" nations or use these forecasts for betting, recall that the history of the Olympic Games is full of surprising performances by individual athletes. No matter whether they were the result of individual excellence or luck, a certain level of unpredictability remains despite all persistence of success. As such, the suspense will remain once again when the world watches the 2016 Olympic Games in Rio.

---

[1] Sponsorship intelligence (2012) London 2012 Olympic Games Global Broadcasting report, http://www.olympic.org/Documents/IOC_Marketing/Broadcasting/London_2012_Global_%20Broadcast_Report.pdf, accessed 28th April 2016.

[2] 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.

[3] 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.

[4] Bredtmann, J., Crede, C. J., and Otten, S. (2016) Participation and Success at the Olympic Games – The Role of Gender Equality. *Mimeo*.

[5] Predictions for all countries can be found at https://sites.google.com/site/olympicmedalspredictions.

[6] This rests on the presumption that no country is excluded from participating at any events due to doping concerns.

[7] Comitê Olímpico do Brasil (COB), http://www.cob.org.br/en/team-brazil/road-to-rio-2016, accessed 13th May 2016.

[8] Wikipedia, https://en.wikipedia.org/wiki/Brazil_at_the_2012_Summer_Olympics, accessed 28th April 2016.