



Munich Personal RePEc Archive

Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometrics Approach

Pfau, Wade Donald

National Graduate Institute for Policy Studies, Japan

December 2006

Online at <https://mpra.ub.uni-muenchen.de/18829/>

MPRA Paper No. 18829, posted 08 Dec 2009 07:03 UTC

This article was published as:

Pfau, W. D., "Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometrics Approach," *The Korean Economic Review*. Vol. 22, No. 2 (Winter 2006), p. 233-247.

***"Predicting the Medal Wins by Country at the 2006 Winter Olympic Games:
An Econometrics Approach"***

by

Wade D. Pfau
National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 162-8677

Phone: 81-90-9300-2243
Email: wpfau@grips.ac.jp

Abstract

Demographic and economic characteristics have been shown to provide important predictive power for determining a country's success in the Olympic Games. This paper extends such research, providing a set of predictions for the gold medals and total medals each country will win at the 2006 Winter Olympics. We expected Germany to win the most medals, followed by the United States, Norway, Italy, Austria, and Canada. For total medals, the overall correlation between the predictions and the actual results was 0.934. While Germany and the United States did finish in the top two places, there were some surprises as Canada, Austria, and Russia performed better than expected, while Norway and Italy did not live up to expectations.

* I wish to provide special thanks to Michael Salmon of the Amateur Athletic Foundation of Los Angeles, as well as to the librarian at National Agency for the Advancement of Sports and Health in Tokyo for their help with accumulating the Olympics participation data.

The 2006 Winter Olympic Games took place during February in Turin, Italy. The Organizing Committee subsequently reported that 2,633 athletes from 80 National Olympic Committees (including first time visits by Albania, Ethiopia, Madagascar, and Serbia and Montenegro) competed in 84 events from seven different sports for 252 medals. This continues an upward growth trend for the Winter Olympics, as there were 78 events and 77 nations represented in 2002. As always, the activities of many fans during the Games included a daily check of national medal standings, which rank participating countries based either on the total number of medals won or the number of gold medals won. Such rankings serve no official purpose, but nonetheless provide enthusiasm (or frustration) to Olympic fans throughout the world. In this paper, we use the methods of economics and econometrics to provide a forecast for the end results of these medal standings before the Olympics took place, in order to demonstrate the power of a simple econometrics model.

Sportswriters make their predictions for Olympics by studying the expected competitors and making a sport-by-sport prediction of who will win each event. Following the theoretical foundation provided in Bernard and Busse (2004), we do not look at these individual competitors. Instead, we rely on the broad economic and demographic characteristics, as well as the past performance, of participating countries to determine how many medals they will win. Undoubtedly, the Olympic Games are about the quests and achievements of individual athletes, and so the sportswriter provides a more invigorating approach to prediction making. However, while specific knowledge about the potential competitors may provide a better prediction for individual events,

adding up the predictions on a sport-by-sport basis can produce a case where the whole is different from the sum of its parts.

This is because there are specific risks, errors and surprises that can happen in each sporting event. One athlete may become injured or have a bad day, while another finds particular inspiration and achieves the performance of a lifetime. Fractions of a second may determine the difference between first place and second place, or between the bronze medalist and a fourth place finish. Judging quirks or other controversies may also arise, which can lead to different results than could have otherwise been expected. Each of these possible events introduces uncertain outcomes into the sportswriters' predictions. The trouble develops when trying to add the predictions together, because we completely lose track of these uncertainties and may produce a medal total that is not really the appropriate sum of the underlying probability distributions for each event's outcome.

The econometric approach used here does not look at individual events, but aims to predict the number of medals won by using broad economic and demographic trends for the participating countries. The actual number of medals won by each country in the previous Olympics provides us with the results after averaging out all of the past uncertainties, which provides historical examples of how these idiosyncrasies came into play. The model is based on the simple idea that larger populations provide a wider pool of potential athletes and that greater economic strength will help to provide the resources for sports training and success. We fit this model using data from past Olympiads, which then allows us to forecast the medal standings for future Olympics.

But again, the Olympics are about individual athletes striving to do their best, which means that our approach will not be perfect. For instance, at the 2002 Winter Olympics, Janica Kostelic of Croatia won 3 gold medals and 1 silver medal in Alpine Skiing events. These were the only medals won by Croatia in their four Olympiads of participation, and anyone using an econometric based model would not have predicted such a stunning performance. Nonetheless, we attempt to demonstrate that a simple econometrics model can produce just as good, if not better, predictions than a sports expert who makes predictions on an event-by-event basis by allowing a more suitable way to “average out” the uncertainty.

As for our findings, we predict that Germany will win the most total medals (35) and gold medals (11) in Turin. In terms of total medals, we predict that the United States will be second, with 31 total medals, and Norway will be third, with 24 total medals. In terms of gold medals though, Norway will tie Germany with 11, while the United States finishes in third with 9. As the host country, Italy should enjoy its most successful Olympic experience since Lillehammer with 20 total medals, 6 of which are gold. Finally, 33 countries should finish these Olympics with a medal, which would be remarkably inclusive considering that previously only 42 countries have ever won at least one medal in the history of the Winter Olympics. Despite a few surprises, these predictions stand up well, as the overall correlation coefficient for total medals was 0.934 and for gold medals was 0.773.

Background to the Study

Several papers proceed as intellectual foundations for this study. Early contributions to the literature include Ball (1972), Grimes et al. (1974), and Levine

(1974). Oddly, research on the topic would then lie dormant for almost 30 years.

Bernard and Busse (2004) helped to revive the literature by developing a theoretical and empirical foundation for predicting the number of medals won by a country. Their model motivates this paper, though their research only considers the Summer Olympics. Their simple model shows that increased populations and greater economic resources should lead to a greater number of medals, because of the larger talent pool and the improved health and sports infrastructure. They also find that host countries generally win more medals than otherwise, that Soviet and planned economies perform better, and that the time needed to build a sports infrastructure is important and observed through lagged medal wins.

Another paper that is closer in spirit to the 1970s research, and which includes results for the Winter Olympics is Johnson and Ali (2004). They seek to determine the structural influences of a country's participation and success. For the Winter Olympics, they find an important role for Gross Domestic Product (GDP) per capita, for the host country status, for single-party or communist systems, and for a heavy winter climate. Unlike the Summer Olympics, they find that population does not play a significant role in explaining medal success, and that smaller countries may even have a slight, though not statistically significant, edge.

Kuper and Sterken (2001) also provide a methodology and forecasts for medal wins and participation at the 2002 Winter Olympics. Their method is to model participation and medal wins as the share of total participation and total medals available. Their forecasts for medal wins are produced with their estimated participation forecasts,

as well as a variable for host country status, income per capita, and a variety of dummy variables for different subsets of countries based on their legal systems.

Data and Methodology

Like the previous literature, we use a regression model to estimate the number of medals a country can be expected to win based on important underlying characteristics. Our model is based on the methodology of Bernard and Busse (2004), which emphasizes a country's population and GDP as key inputs into the production of medal wins at the Olympics. If sporting talent is distributed randomly throughout the world and can be enhanced by training in world-class sporting facilities, then larger populations and richer economies should both help produce better athletes capable of winning Olympic medals. In this production process, other factors, such as serving as the host country or having a Communist government, may help to produce a greater output given a particular level of demographic and economic inputs. Also, past success at the Olympics can help explain present successes by indicating that a sports infrastructure is in place in the country.

As such, our dependent variable is the share of medals (*MS*) won by a country at a given Olympics. For explanatory variables, we expect that countries with larger populations should have an edge in drawing from a larger pool of potential talent. Countries with greater economic resources also have the potential to provide better training conditions for their athletes. Nonetheless, there are constraints on a country's total participation, so these influences should not grow linearly without bound. We must also take care to account for correlation between a country's population and total GDP by examining the GDP per capita. Thus, we include the logarithms of population (*POP*) and of GDP per capita (*GDPCAP*).

Additionally, host countries (*HOST*) can be expected to perform better on account of the increased audience support, familiarity with the sporting facilities, and other factors. It is also reasonable to expect some inertia in medal winning performances, and so we include the lagged share of medal wins for a country. Certain subsets of countries have also performed unusually well given their circumstances and have historically accounted for the vast majority of medal wins, warranting the inclusion of additional dummy variables, all of which we expect to have positive coefficients. These include a Soviet dummy variable (*SOV*) for Bulgaria, East Germany, Hungary, Poland, Romania, Czechoslovakia, and the Soviet Union until 1988. The Scandanavia variable (*SCAND*) includes Norway, Sweden, and Finland. The Germanic country variable (*GERM*) includes Germany, Austria, West Germany, and East Germany. The Alpine country variable (*ALPINE*) includes Switzerland, Italy, and France, and the North America variable (*NORTHAM*) includes the United States and Canada. The regression looks as follows:

$$MS_t = \beta_0 + \beta_1 \log(\text{POP}_t) + \beta_2 \log(\text{GDPCAP}_t) + \beta_3 \text{HOST}_t + \beta_4 MS_{t-1} + \beta_5 \text{SOV}_t + \beta_6 \text{SCAND}_t + \beta_7 \text{GERM}_t + \beta_8 \text{ALPINE}_t + \beta_9 \text{NORTHAM}_t + \varepsilon_t$$

This model is similar in spirit to Bernard and Busse (2004), though their predictions are only for the Summer Olympics. With respect to the other papers we cited, neither Kuper and Sterken (2001) or Johnson and Ali (2004) include an explanatory variable for lagged medal totals, thus missing the opportunity to have such a proxy that measures existing sports infrastructure. Instead, Kuper and Sterken (2001) use the share of participants competing for each country, which in effect can serve to some extent as a proxy for existing sports infrastructure as well. Johnson and Ali (2004) also use the total medals won instead of the share of medals won, and they include a time trend to allow for

the fact that the total medals available is increasing over time. Because the total available medals do not grow at a constant pace, we question whether this gives a better specification than using medal shares. With regard to which model is best, it is almost certainly too soon to tell though. Though one model may perform better than others for one particular Olympiad, the outcome could be reversed for the next Olympiad. The model here contributes by exploring another specification to see how well a simple econometrics model can forecast Olympic results.

When making the out-of-sample predictions for 2002, the data for the Winter Olympics from 1960 to 1998 is used. To extrapolate after 1998, we assume the same list of participating nations in 2002 as in 1998, and we assume the same populations and GDPs in 2002 as in 1998. This is necessary because the more recent data would not have been available prior to the Olympiad. Likewise for the 2006 out-of-sample predictions, we use the data from 1960 to 2002, and assume that the list of participating nations in 2006 is the same as 2002.

The lists of nations participating at each Olympiad were collected from several sources, including the *Official Reports* for various Olympiads and a computer spreadsheet provided by the Amateur Athletic Foundation of Los Angeles, which cites Lyberg (1994) and Gueorguiev (1998). Data on medal wins for each country at each Olympiad comes from the International Olympic Committee website. We also consulted the *Official Reports* for each Olympiad and Wallechinsky (2001), and we found occasional discrepancies, so we assume that the International Olympic Committee website provides the most recent medal updates after accounting for later disqualifications and other matters. Data for population and GDP measured in constant 1995 US dollars came

primarily from the World Bank's World Development Indicators 2004 CD-ROM.

Occasionally, it was necessary to extrapolate data or use other sources for countries not included in the World Bank database.

Results

Our results include the fitted regression model, a comparison of various out-of-sample predictions made for the 2002 Winter Olympics, and our own predictions for gold medals and total medals won by each country at the 2006 Winter Olympics. Table 1 shows four sets of regression results. These include estimations for total medals won and gold medals won from 1960 through 2002 (for the 2006 predictions) and from 1960 to 1998 (for the 2002 predictions). As expected, all of the explanatory variables have positive coefficients, though statistical significance at the 5% level is not quite found for population or income per capita. The only exception is that population helps to explain the share of gold medals won. Regarding population, unlike the Summer Olympics, population may be less important because skills for winter sports are specialized to colder regions of the world, particularly Scandinavia, which tend to have smaller populations. The reason that income per capita is not significant can be explained by the inclusion of the lagged medal totals and country specific dummy variables. Without these variables, we found that income per capita is highly significant, but the inclusion of these other explanatory variables means that there is little else for income per capita to explain. Meanwhile, serving as the host city does have an important advantage for medal winning, and the host city effect for the 2006 Olympics should translate into an additional 7 medals for Italy than it would have otherwise won. The Soviet, Scandinavia, and Germanic dummy variables are also highly significant, with these countries winning

more medals than would otherwise be justified by their economic and demographic characteristics.

Tables 2 and 3 compare the out-of-sample forecasts this approach would have provided for the 2002 Winter Olympics to the forecasts made by Kuper and Sterken (2001) and Johnson and Ali (2004). For Kuper and Sterken (2001), we use the forecasts from their Table 3, which are created using “country specific intercepts.” All three approaches produced fairly accurate predictions. Some noticeable exceptions include the surprisingly strong performances of Germany and the United States, and the performances of Russia, Finland, and Japan falling below the forecasters’ expectations. In terms of performance, the forecasts of this paper for total medals come out slightly ahead of the others in terms of the correlation coefficients. This paper’s approach was noticeably more successful than the others in producing forecasts for gold medals. The Johnson and Ali (2004) forecasts were the least successful, though the authors do indicate that their goal was to explain important influences on the medal results rather than producing the most accurate possible predictions.

This leads us to the heart of this paper, which consists of the forecasts for the 2006 Winter Olympics and their comparison to the actual results, provided in Tables 4 and 5. For total medals, the top three rankings should match the 2002 Winter Olympics, as Germany will come in first place with 35 medals, the United States in second place with 31 medals, and Norway in third place with 24 medals. The United States medal count should fall from the 34 medals won in Salt Lake City (the United States Olympic Committee was hoping for 20 medals at those Olympics, which would still have been a large improvement from the 13 medals won in 1998) because it is no longer the host

country. As host country this year, Italy should surpass Austria and Canada to come in fourth place with 20 total medals. Austria follows with 19 medals, Canada with 17 medals, France with 12 medals, and Russia and Switzerland with 11 medals each. Finland and Sweden should round-out the top 10 rankings with 10 medals each. Argentina, Brazil, Chinese Taipei, Hong Kong, India, and Mexico are all poised to win their first medal in the Winter Olympics. Table 5 continues with the forecasts specifically for gold medals. Here, Germany and Norway should tie with 11 gold medals each. Next, the United States can be expected to win 9 gold medals, while Canada wins 7 gold medals, and Italy wins 6 gold medals.

Regarding how these econometrics based forecasts fared, the overall correlation coefficient for total medals was 0.934. As expected, Germany and the United States finished at the top of the rankings, though each won 6 less medals than predicted. Then, there were some surprises. Canada finished third in the rankings, with 7 more medals than expected, followed by Austria with 4 more medals than expected. Fifth was Russia, which earned 22 medals, 11 more medals than expected. Other countries which won more medals than expected include South Korea, Sweden, China, and Switzerland. In South Korea's case, for example, the 11 medals won exceeds the 4 medals won in 2002 and the country's previous record of 6 medals won in both 1994 and 1998.

On the other hand, some countries did not perform as well as expected. In particular, the host advantage for Italy led us to expect 20 medals, while Italy finished with only 9 medals. Norway also had a disappointing performance, as we expected the country to finish third with 24 medals, but at the end, Norway had only 19 medals. At the other end of the spectrum, our prediction that this Olympics could be the most

inclusive of all did not happen, as of the 11 countries we expected to win one medal, only Belarus accomplished this goal.

In terms of gold medals, the overall correlation coefficient was 0.773, and we accurately predicted 11 golds for Germany, 9 golds for the United States, and 7 golds for Canada. Near the top of the rankings, there were other surprises as well. Austria tied the United States with 9 gold medals, and Sweden won 7 gold medals, which was 5 more than expected in both cases. Russia and South Korea also both won 4 more gold medals than expected. Meanwhile, while Sweden was a relative over-performer in Scandinavia, both Norway and Finland both significantly underperformed. Norway finished with 2 gold medals, 9 less than expected, and Finland finished with 0 gold medals, 5 less than expected.

Conclusion

Past research has shown that the demographic and economic characteristics of a country have important predictive powers for determining how many medals a country can be expected to win in the Olympic Games. This paper builds a model along these lines in order to provide predictions for the 2006 Winter Olympics in Turin. The model performs fairly well, but still leaves plenty of room for improvements in the future. First, effort should be made to account for the noisiness of the results in the Winter Olympics, as the medal totals can fluctuate by a large degree between Olympiads. For example, Norway won 5 medals in 1988 and 20 medals in 1992, the United States went from 13 medals in 1998 to 34 medals in 2002 to 25 medals in 2006, Austria had 21 medals in 1992, 9 medals in 1994, and 17 medals in 1998, and so on. The inconsistency of country performances makes the use of simple econometric models more challenging. A

potential avenue to explore for improving the model prediction would be to incorporate the results of international sporting events leading up to the Olympics as explanatory variables. This could potentially improve the predictions by adding information about how the situation has changed since the previous Olympics, which would reduce the reliance on the lagged medal totals variable. Finally, efforts should also be made to account for how many athletes are sent by each country. For example, while demographic and economic characteristics led us to predict a medal for Chinese Taipei and Hong Kong, both countries actually only sent one athlete apiece, information which was not accounted for in the current form of the model.

Bibliography

- Ball, Donald W. "Olympic Games Competition: Structural Correlates of National Success." *International Journal of Comparative Sociology* Vol. 12 (1972): 186-200.
- Bernard, Andrew B., and Meghan R. Busse. "Who Wins the Olympic Games: Economic Resources and Medal Totals." *Review of Economics and Statistics* Vol. 86, No. 1 (February 2004): 413-417.
- International Olympic Committee. *Official Report of the Games*. Lausanne, Switzerland: International Olympic Committee, Various Years.
- Grimes, A. Ray, William J. Kelly, and Paul H. Rubin. "A Socioeconomic Model of National Olympic Performance." *Social Science Quarterly* Vol. 55 (1974): 777-782.
- Gueorguiev, N. *La Participation nationale aux Jeux Olympiques d'hiver 1924-1998*. 1998.
- Johnson, Daniel K. N., and Ayfer Ali. "A Tale of Two Seasons: Participation and Medal Counts at the Summer and Winter Olympic Games." *Social Science Quarterly* Vol. 85, No. 4 (December 2004): 974-993.
- Kuper, Gerard, and Elmer Sterken. "The Olympic Winter Games: Participation and Performance." University of Groningen, mimeo, December 2001.
- Levine, Ned. "Why Do Countries Win Olympic Medals? Some Structural Correlates of Olympic Games Success: 1972." *Sociology and Social Research* Vol. 58 (1974): 353-360.
- Lyberg, Wolf. *The Book of Facts on the Olympic Winter Games, 1908-1994*. 1994.
- Wallechinsky, David. *The Complete Book of the Winter Olympics: Salt Lake City 2002 Edition*. Woodstock, NY: The Overlook Press, 2001.
- World Bank. *World Development Indicators 2004 CD-ROM*. April 2004.

Table 1:
Share of Medals Won at the Winter Olympic Games

Variable	Estimation Covering 1960 - 2002		Estimation Covering 1960 - 1998	
	Share of Total Medals	Share of Gold Medals	Share of Total Medals	Share of Gold Medals
log(Pop)	0.0007	0.0011*	0.0008	0.0013*
log(GDP/capita)	0.0009	0.0015	0.0011	0.0019
Host Country Lagged Medal Share	.0277***	0.0309***	.0231***	0.0298***
Soviet Union	0.7410***	0.6717***	0.7409***	0.6728***
Scandinavia	0.0124***	0.0165***	0.0123***	0.0164***
Germany	0.0128***	0.0175**	0.0132***	0.0159**
Alpine	0.0194***	0.0215***	0.0187***	0.0212***
North America	0.0072	0.0068	0.006	0.0045
Constant	0.0102*	0.0136*	0.0073	0.0114
	-0.0082	-0.0142	-0.0104	-0.0177

Note: * 5% significance, ** 1% significance, *** 0.1% significance

**Table 2:
Comparison of Out-of-Sample Forecasts
for the 2002 Winter Olympics (Total Medals)**

Country Name	Total Medals	Kuper and Sterken	Johnson and Ali	Pfau
Germany	36	27	31	30
United States	34	25	20	20
Norway	25	27	20	25
Canada	17	14	11	15
Austria	17	17	16	20
Russian Federation	13	24	21	16
Italy	13	15	11	11
France	11	8	8	9
Switzerland	11	11	13	8
Netherlands	8	7	13	10
China	8	5	9	7
Finland	7	15	14	14
Sweden	7	7	10	6
Croatia	4	0	0	0
Korea	4	6	4	6
Czech Republic	3	3	4	3
Estonia	3	0	0	0
Bulgaria	3	0	0	1
Australia	2	1	3	2
Great Britain	2	1	4	2
Japan	2	6	7	10
Poland	2	0	0	0
Belarus	1	3	0	2
Slovenia	1	2	0	0
Kazakhstan	0	3	0	2
Ukraine	0	1	0	1
Belgium	0	0	3	2
Denmark	0	0	3	2
Brazil	0	0	0	1
Chinese Taipei	0	0	0	1
Spain	0	0	0	1
Argentina	0	0	0	1
Iceland	0	0	4	0
Luxembourg	0	0	3	0
Total Medals	234	228	232	228
Correlation Coefficient*		0.938	0.925	0.942

* Correlation coefficient is for the 77 participating countries

Table 3:
Comparison of Out-of-Sample Forecasts
for the 2002 Winter Olympics (Gold Medals)

Country Name	Gold Medals	Kuper and Sterken	Johnson and Ali	Pfau
Norway	13	7	6	9
Germany	12	10	11	12
United States	10	14	7	9
Canada	7	2	3	6
Russian Federation	5	11	10	7
France	4	1	2	2
Italy	4	8	3	2
Finland	4	4	4	3
Netherlands	3	2	6	4
Austria	3	4	4	4
Switzerland	3	4	4	2
Croatia	3	0	0	0
China	2	0	0	0
Korea	2	2	2	3
Australia	2	0	1	0
Czech Republic	1	0	1	1
Estonia	1	0	0	0
Great Britain	1	1	1	1
Sweden	0	4	3	2
Japan	0	1	6	5
Iceland	0	0	1	0
Belgium	0	0	1	0
Luxembourg	0	0	1	0
Denmark	0	0	1	0
Bulgaria	0	0	0	1
Total Gold Medals	80	75	78	73
Correlation Coefficient		0.826	0.805	0.914

* Correlation coefficient is for the 77 participating countries

Table 4:
Results for the 2006 Winter Olympics (Total Medals)

Country Name	Total Medals	2006 Total Medals Prediction	Difference
Germany	29	35	-6
United States	25	31	-6
Canada	24	17	7
Austria	23	19	4
Russian Federation	22	11	11
Norway	19	24	-5
Sweden	14	10	4
Switzerland	14	11	3
China	11	7	4
Italy	11	20	-9
South Korea	11	4	7
Finland	9	10	-1
France	9	12	-3
Netherlands	9	7	2
Czech Republic	4	3	1
Croatia	3	3	0
Estonia	3	2	1
Australia	2	2	0
Poland	2	2	0
Belarus	1	1	0
Bulgaria	1	2	-1
Great Britain	1	3	-2
Japan	1	3	-2
Latvia	1	0	1
Slovakia	1	0	1
Ukraine	1	0	1
Argentina	0	1	-1
Belgium	0	1	-1
Brazil	0	1	-1
Chinese Taipei	0	1	-1
Denmark	0	1	-1
Hong Kong	0	1	-1
India	0	1	-1
Mexico	0	1	-1
Slovenia	0	1	-1
Spain	0	1	-1
Total Medals	251	249	
Correlation Coefficient*		0.934	

* Correlation coefficient is for the 80 participating countries

**Table 5:
Results for the 2006 Winter Olympics (Gold Medals)**

Country Name	2006 Actual Gold Medals	2006 Gold Medals Prediction	Difference
Germany	11	11	0
United States	9	9	0
Austria	9	4	5
Russian Federation	8	4	4
Canada	7	7	0
Sweden	7	2	5
South Korea	6	2	4
Italy	5	6	-1
Switzerland	5	3	2
Estonia	3	1	2
France	3	4	-1
Netherlands	3	3	0
China	2	2	0
Norway	2	11	-9
Australia	1	2	-1
Croatia	1	2	-1
Czech Republic	1	1	0
Japan	1	1	0
Finland	0	5	-5
Great Britain	0	1	-1
Total Medals	84	81	
Correlation Coefficient*		0.773	

* Correlation coefficient is for the 80 participating countries