The Miracle on Thin Ice: How A Nation's GDP Affects its Olympic Performance

Jennifer Boudreau, James Kepner, and Joseph Rondone

The Miracle on Thin Ice: How A Nation's GDP Affects its Olympic Performance

Abstract

The Olympics offer participating countries the opportunity to promote their best athletes and compete with other countries. In this paper, we compare and test country statistics in order to determine what factors have a large impact on their Olympic performance. We estimate that GDP per capita is the main determinant to a country's successful Olympic performance, quantified by amount of gold, silver, and bronze medals a country receives. However, our results indicate that country size and health care expenditure per capita were significant in determining Olympic performance as opposed to GDP per capita, disproving our initial hypothesis.

Keywords:
Olympics, Sports, Economics, Regression, Econometrics

I. Introduction

The Olympics is an exciting time for the world. They allow a host nation to show off their best for the world. Alternatively, the Olympics can form a platform for people of all nations to interact and mingle without the need for geopolitical tensions. The Olympics provide a platform for the world to come together and celebrate something everyone loves: sports. Sports are a platform which can be used to unify people. Sports are often seen as a place where individuals, communities, and whole nations can become inspired. This is the reason why the Olympics, and major sporting events like it, are watched by hundreds of millions of people worldwide.

This research project will analyze how various countries do at the Olympics. We believe that due to Olympic performance being a source of national pride, how a country performs is important for that country's global perception. As such, countries have a large incentive to perform well at the Olympics. This paper will look into the relationship between a country's performance at the Olympics and the size of its economy. We believe that as a country's gross domestic product (GDP) grows, it will perform better at the Olympics. However, Olympic performance is not tied solely to a country's GDP. Therefore, we want to test Olympic performance with a myriad of other factors to see if there is any ensuing change. By conducting this research, we further not only our knowledge of Olympians but also how that relates to the geopolitical structure of the world.

II. Literature Review

For many decades, the Olympics have not only been a time for countries to come together and celebrate peace and sportsmanship, but also an opportunity to display their prowess and abilities in the realm of athletics. The factors that contribute to high medal counts seen in countries such as the United States of America, Russia and the United Kingdom have been widely studied by political scientists and economists alike. There are a multitude of factors that can contribute to a country's success as the Olympics that are quantifiable, such as Gross Domestic Product (GDP) and population, as well as factors that are much harder to capture and quantify, such as cultural and political factors. Throughout the years, multiple studies have been conducted to conclude if any factor can truly explain what factors, if any, can truly be the sources of certain countries' presumed athletic superiority.

Robert Hoffman et al. (2002) conducted an interesting study regarding Olympic success that looked at multiple variables: GNP per capita, population, geographical factors (using average climate and rainfall for each capital in the respective country), sports funding derived from GNP, political factors and cultural factors. The two interesting aspects of this study as compared to other studies is as follows: for GNP per capita and population they were subject to the power of ½ to account for diminishing returns as increase in wealth and population does not account for the same increase in resources and talent pool. In

addition to accounting for diminishing returns, this paper also takes into account climate factors. Interestingly, countries with an average temperature of 16 degrees Celsius averaged around 15 medals suggesting that countries with mild average climate have a more advantageous environment for nurturing athletes. (However, it should be noted that this should be taken with a grain of salt—countries like Russia, China, and the United States, all huge medal contenders each Olympics, have vast swaths of land that cover a variety of different biospheres and climates.)

Moosa and Smith (2004) go another route to discuss factors that affect Olympic success. While Hoffman found that factors such as GNP and population had the biggest effect on Olympic success, there are still huge anomalies in the data. For example, in the 1996 Atlanta Olympics, Cuba, a country a fraction of the size of India won 25 medals. India on the other hand? The country only succeeded in winning one medal. In the 2000 Olympics Cuba won 29 medals while India won, again, one medal. Seeing as India surpasses Cuba in GNI and population by a very large margin, how can this be explained? There are a few interesting points to take note of from this study. For example, on the note of population, while Vietnam and Germany both had sizeable populations, the former sent only 7 athletes in the 1996 Olympics while the latter sent over 400 athletes in a variety of different events. That same year 1.5 billion of the world's poorest people won only 3% of all medals. From this study it shows that GDP, population, number of athletes presenting a country and health expenditure are instrumental in determining how successful a country's attempt at winning the most Olympic medals as they are able to determine most accurately the size of the economy and the resources available to the sports teams and training. However, the extreme bound model used still fails to explain outliers such as Australia, East Germany, and Cuba.

In a different study, Lui and Suen (2008) take a slightly different approach. Instead of looking at just the medal count data from one Olympic game, it combines all the data from the years 1952-2004 games. One of the interesting things pointed out by this article is that often times not all top athletes can compete for their country based on ceilings set by the International Olympics Committee (IOC). Thus, it is not uncommon to see world-class athletes renounce their nationality in favor of competing for a different country. While most of the variables are standard such as income and population, this study introduced a covariate seen in very few or any other studies: education. While education seems at first thought an interesting variable to consider, there was no proven bearing on the number of medals won. However, as the data obtained pertained to education below the collegiate level, it may in the future prove worth a try to see if collegiate sports program on a national level help to promote Olympians.

This research conducted tries to provide a new perspective into the realm of Olympic superiority. Of much of the literature found thus far, little of the analysis has been done post the 2000 Olympics. Analyzing the Olympics from 2004 and 2008, both landmarks in their own right, might provide some more specific insight as to what factors are big predictors of Olympic medal counts. Also, instead of

looking at climate temperatures, this research chose to use the central latitude of a country to see if that can bring out a stronger correlation between the number of medals won. If one is considering geographic factors, it is also possible to argue that countries with a higher concentration of people living in urban areas will have higher Olympic success; living in urban areas provide budding athletes with better facilities and access to trainers and resources. Using proxies for models can often be seen as misleading as factors such as culture and politics are very hard to quantify, so the approach taken is to try and identify strictly quantitative factors that affect Olympic medal outcomes. In an attempt to gather more comprehensive analysis in this study the medals are weighted as opposed to not being weighted: gold medals count as six, silver as four, and bronze as two.

III. Data

This study is primarily focused on the impact of country's GDP on its Olympic performance. The dependent variable of this study is a country's Olympic performance. We measured this by recording all medal recipients for the 2004 and 2008 Summer Olympics by country. We then weighted the medal using a point system (6 for gold, 4 for silver, and 2 for bronze) to adjust for the value of each particular medal. Therefore, the country with the most points "wins" the Olympics.

In addition to the dependent variable, we incorporated independent variables into the study. First and foremost is a country's GDP. We took GDP data from the World Bank and computed the Log (e) of that data. Log (e) made the data a more manageable size without compromising the integrity of that data. GDP is the primary independent variable because it has the greatest ability to affect Olympic performance. Countries with a high GDP are more likely to have the resources to build training facilities and support an active sports community than those with lower GDPs.

Beyond a country's GDP, we also took a look at five other variables which might have an impact on Olympic performance. First is the population of a country. Countries with larger populations have a large pool in which athletes can come from, having a potentially high impact on Olympic performance. Second is the urban percentage of the country. Countries which are heavily urbanized might have better facilities to identify and foster athletes. Additionally, urban communities have a greater ability to support and lobby for sports venues than rural communities. Thus there are access differences between urban and rural communities which might impact Olympic performance. Third is the nature of a country's economy. We identified OECD nations as being the generally more advanced and industrial economies than non-OECD countries. Therefore, we incorporated a binomial variable for OECD and non-OECD countries in our study. Fourth is per capita health expenditure for a country. Countries which spend more on health might have a greater ability to support athletes than those who do not. Additionally, healthier countries can potentially provide better quality athletes than those who are not. The last variable we looked at was

the latitude of the country. As a country moves further from the equator, there is a notable difference in the country make up, a phenomenon often called the North-South divide. This can potentially impact a country's performance and ability to compete in the Olympics.

These six independent variables provide us with a foundation upon which we can build our statistical analysis. As we will see, the interaction of these variables will hopefully provide a better picture into the causes and influencers of Olympic performance.

2. Source of Data

There are two primary sources of data used in this study. The first comes from the CIA World Factbook. This source provided us with the demographic and geographic information required by the study. It is one of the official clearinghouses of government data on other countries. Additionally, the site is updated continuously by the U.S. Federal government. The second source of data was from the World Bank. This source provided us with the economic data used in the study such as GDP and per capita health expenditure. The World Bank produces some of the most reliable economic information on countries and therefore is a realizable source for this study.

Labels	**Variable**
totalscore	Weighted medal count: Gold -6, Silver – 4, Bronze - 2
new_gdp	Log (per capita GDP)
oecdcountry	Binary variable: yes or no to OECD membership
latitude	Latitude of country's capital
healthexpe~e	Healthcare expenditure per capita
urban	Percentage urban population
avgage	Average age
newarea	Log (country's area in km^2)

3. Descriptive Statistics

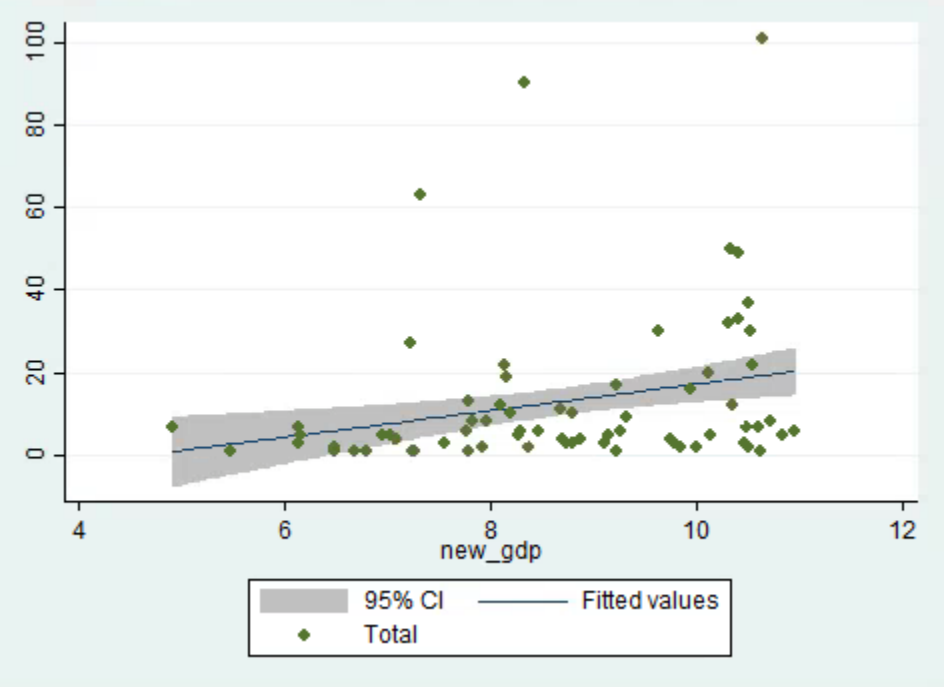
Based on the raw data we have gathered, we decided to create three new variables: new_gdp, new_pop, and total score. New_gdp was found by taking the natural log of GDP per capita in order to reduce the effect that such large numbers would have on our regression. New_pop was found by taking the natural log of the country's population and was done for the same reason as new_gdp. Total score was calculated by multiplying the number of gold medals a country received by 6, silver medals by 4, and bronze medals by 2. This was done in order to give more weight to the number gold medals a country received; the multiples of 2 were chosen in order to increase the spread of the data.

Simple regression model:

$y = \beta_0 + 12.929\,\beta_1$

Multiple regression model:

$y = \beta_0 + 15.778\,\beta_1 - 30.969\,\beta_2 + 0.0085\,\beta_3 + 0.013\,\beta_4 - 0.730\,\beta_5 + 1.311\,\beta_6 + 24.860\,\beta_7$



4. Check whether the data meets the Gauss Markov Assumptions

This data set and its variables meet the Gauss Markov Assumptions. This data in linear in parameter, increasing or decreasing in a predictable manner, and no variable is not squared in the

equation, making it a linear function. The sample data is random, with every data point being equally likely to be picked as the other. To come close to conditional mean of error even if we know one variable, we cannot predict if they are above the population regression line. Furthermore, no variables are perfectly collinear, even though many variables like OECD membership and per capita GDP are highly and positively correlated. While it is difficult to check for homoscedastic errors, the presence of an abundance of independent variables will mitigate any unknown homoscedastic error. Finally, there is no serial correlation between errors. Therefore, the least squared estimates have no other linear unbiased estimates, making them the best linear unbiased estimate.

	totals~e	new_gdp	oecdco~y	latitude	health~e	urban	avgage	newarea
totalscore	1.0000							
new_gdp	0.2446	1.0000						
oecdcountry	0.1522	0.7429	1.0000					
latitude	0.1253	0.3135	0.3951	1.0000				
healthexpe~e	0.3582	0.7927	0.6793	0.2920	1.0000			
urban	0.1940	0.6985	0.4711	0.1896	0.4909	1.0000		
avgage	0.2111	0.7823	0.6378	0.3634	0.5944	0.6512	1.0000	
newarea	0.4457	-0.2456	-0.1926	-0.2316	-0.0723	0.0290	-0.2098	1.0000

III. Results

1. Provide STATA results (in table form) of the estimated equation: Simple Regression

Source	SS	df	MS		
Model	53679.6361	1	53679.6361	Prob > F	0.0027
Residual	795015.335	138	5760.98069	R-squared	0.0632
				Adj R-squared	0.0565
Total	848694.971	139	6105.71922	Root MSE	75.901

Number of obs	=	140				
F(1, 138)	=	9.32				
Prob > F	=	0.0027				
R-squared	=	0.0632				
Adj R-squared	=	0.0565				
Root MSE	=	75.901				

totalscore	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
new_gdp	12.92947	4.235691	3.05	0.003	4.554227	21.30472
_cons	-60.63321	37.2953	-1.63	0.106	-134.3773	13.11091

In the initial simple regression, the relationship between a country's performance at the Olympics and their per capita GDP was proved to be statistically significant. This confirms our hypothesis that there per capita GDP impacts performance. Looking at the t-statistic, we can see that this relationship is

significant at the 1% level. Moreover, the relationship is a positive sloping relationship, showing that GDP improves Olympic performance. In order to dig into the relationship, a multiple regression analysis is required.

2. Provide STATA results (in table form) of the estimated equation: Multiple Regression

Source	SS	df	MS		Number of obs =	138
					F(7, 130) =	12.46
Model	339582.393	7	48511.7705		Prob > F =	0.0000
Residual	505944.563	130	3891.88126		R-squared =	0.4016
					Adj R-squared =	0.3694
Total	845526.957	137	6171.72961		Root MSE =	62.385

totalscore	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
new_gdp	15.77754	9.173139	1.72	0.088	-2.37042	33.9255
oecdcountry	-30.96865	17.1086	-1.81	0.073	-64.81596	2.878661
latitude	.0085103	.0042856	1.99	0.049	.0000318	.0169888
healthexpe~e	.0130791	.0059478	2.20	0.030	.0013121	.0248462
urban	-.729927	.4269717	-1.71	0.090	-1.57464	.1147854
avgage	1.311437	1.217285	1.08	0.283	-1.096816	3.719691
newarea	24.86008	3.385513	7.34	0.000	18.16224	31.55791
_cons	-465.2306	97.31306	-4.78	0.000	-657.7528	-272.7083

Unrestricted multiple variable regression model

In the unrestricted multiple variable regression model, we explore seven variables which might influence Olympic performance. Looking at the regression results, only three of the seven variables were significant at the 5% level: capitol city latitude, per capita health expenditure, and the natural log of a country's area. This model disproves our hypothesis that per capita GDP is tied to Olympic performance. Instead, three unexpected variables were significant. The relationship between Olympic performance, area, and per capita health expenditure makes economic sense. If a population spends more on health, the overall well-being of that population is expected to be better. Additionally, that country should have a greater health infrastructure to train and invest in athletes. Alternatively, a larger country by area gives athletes a greater chance to experience a more diverse environment. Moreover, more territory gives the athletes more room to train. Meanwhile, it is interesting that latitude was statistically significant due to the large expanse of major countries. States such as the United States, Russia, and China draw athletes from multiple latitudes. However, this follows our reasoning that countries from more temperate climates perform better at the Olympics.

Source	SS	df	MS		
Model	311651.132	3	103883.711	Number of obs =	138
Residual	533875.825	134	3984.14795	F(3, 134) =	26.07
				Prob > F =	0.0000
Total	845526.957	137	6171.72961	R-squared =	0.3686
				Adj R-squared =	0.3545
				Root MSE =	63.12

totalscore	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
latitude	.0077769	.0041325	1.88	0.062	-.0003964	.0159502
healthexpenditure	.0175837	.0035635	4.93	0.000	.0105357	.0246316
newarea	22.05285	3.091188	7.13	0.000	15.93902	28.16668
_cons	-262.1255	41.79491	-6.27	0.000	-344.7885	-179.4624

Restricted multiple variable regression model

After running the multiple variable regression, we decided to rerun the regression with a restricted variable set. We chose only the variables which were statistically significant in the previous model. By doing this, we hope to focus in on the variables which directly correlate with Olympic performance. Looking at the results, we can see that per capita health expenditure and the natural log of a country's area are the only statistically significant variables at the 5% level. Latitude is statistically significant at the 10% level but barely falls short with a t-test. By running this model, we can see which variables are most important in Olympic performance. As explained above, per capita health expenditure and the size of a country are the two variables which stand the statistical test.

Dependent Variable log (salary)			
Independent Variables	**Simple Regression**	**Unrestricted Multiple Regression**	**Restricted Multiple Regression**
Ln per capita GDP	**12.92947 *** (3.05)**	**15.78 * (1.72)**	
OECD Membership		**-30.97 * (-1.81)**	
Latitude		**0.01 ****	**0.01***

		(1.99)	**(1.88)**
Health care expenditure per capita		**0.01*** (2.20)**	**0.02*** (4.63)**
Average Age		**1.31 (1.08)**	
Ln of State Area, sq. km		**24.86*** (7.34)**	**22.05*** (7.13)**
Intercept	**-60.63* (-1.63)**	**-465.23 *** (-4.78)**	**-262.13*** (-6.27)**
No. of obs.	**140**	**138**	**138**
R-square	**0.3**	**0.4016**	**0.3686**

The F-statistic shows robustness for the model in explaining the relationship. The F-statistic for this model is 2.445. The critical value for this model, however, is 2.6. Therefore, due to this model having a small F-statistic, this model does not have a robust explanatory power. As a result, our variables are not jointly statistically significant despite their individual significance at the 5% level.

3. Interpret the results

As a result of running our simple regression model, we have come across interesting results. As expected, there is a positive correlation between new_gdp and totalscore, meaning that the higher a country's per capita GDP, the better that chance that it will win more medals in the Olympics. This coincides with our hypothesis that richer countries have a greater means to find, develop, and support athletes than poorer countries. In the multiple regression analysis, this trend continued disappeared. In fact, the multiple regression analysis shows that per capita GDP does not statistically affect Olympic performance. Instead, the country's area and per capita health expenditure has a much higher impact on Olympic performance than any other variable. One explanation could be that the greater a country's area is, the more diverse of an environment available for the athletes to train in. Additionally, large countries such as United States, China, Germany, Canada, and Russia historically perform exceptionally well in the Olympics, explaining how the newarea variable has such a high level of significance since these countries

skew the variable upwards. Additionally, health care expenditure per capita and latitude were also statically significant. The higher the per capita health care expenditure is, the healthier a country's population tends to be, increasing their chances of producing and taking care of superb athletes. The latitude of the country's capital impacting Olympic performance can be explained by the high performance of the United States, Russia, China, and other European countries, which have higher latitudes and these performances may have slightly skewed the variable.

On the other hand, a country's membership into the OECD has a negative effect on their Olympic performance. This came as a surprise to us. After reexamining the data, we hypothesize that the exclusion of certain high-performing countries like China and Russia from the OECD has caused this negative correlation. Beyond these, the other three variables (GDP per capita, percent urban population, and average age) had a negligible impact on a country's performance.

After removing these insignificant variables and running the multiple regression model with only latitude, health expenditure, and the log of the country's area, we discover that latitude of the state's capitol becomes less significant in determining a country's Olympic performance compared to the other two variables. Comparing the other two variables, we found that the log of the country's area was the most significant variable in determining the country's Olympic performance. We believe that the reason for this significance is the same as that as listed above.

Conclusion:

Every four years the Olympics draws the eyes of millions to one place, one event. There is no single sporting event that can match the ferocity or the intensity exhibited by these athletes and those who cheer for their countries. Over the course of this research paper we have looked at the following factors: natural log of GDP per capita, natural log of total land mass of a country in km^2, healthcare expenditure per capita, latitude of a country's capital city, whether a country is a member in the OECD country, percentage of urban population as part of the whole, as well as the average age of a country. To create a more fair representation of a country's medal count we made the choice to weight each medal: respectively for gold, silver and bronze the weighted count was six, four and two.

While we initially assumed GDP per capita would be the most significant factor in determining medal count success, the results yielded interesting results. At the one-percent level healthcare expenditure per capita as well as the log of the country's area were found to be significant. At the five-percent level latitude was significant, and then at the ten percent level OECD membership as well as the log of GDP per capita was significant.

Predicting the success of a county at the Olympics is not an exact science, but these findings do lend some insight into this fascinating subject. Healthcare expenditure per capita is significant as the better the general welfare of the country itself the better the health care that is available to the athletes themselves. Large land masses could indicate that countries with a larger area, i.e. more diverse climates and landscapes, could offer better environments for athletic training compared to smaller countries whose geographical boundaries limit their ability to train effectively. GDP per capita may not be a very significant factor in determining success because even in places with lower standards of living the government might take on the bulk of expenses in training their athletes in places such as China or Russia. In a future study, it would be interesting to run this same model excluding data from the P5 + 1 (China, Great Britain, the United States, Russia, France and Germany) to see how the data points fall.

Work Cited

Hoffman R, Ging L, Ramasamy B. Public Policy and Olympic Success. *Applied Economics Letters* [serial online]. June 2002;9(8):545-548. Available from: EconLit, Ipswich, MA. Accessed March 13, 2014.

Lui, H., & Suen, W. (2008). Men, Money, and Medals: An econometric Analysis of the Olympic Games. *Pacific Economic Review*, 13(1), 1-16

Moosa I, Smith L. Economic Development Indicators as Determinants of Medal Winning at the Sydney Olympics: An Extreme Bounds Analysis. *Australian Economic Papers* [serial online]. September 2004;43(3):288-301.Available from: EconLit, Ipswich, MA. Accessed March 13, 2014.

World Development Indicators, The World Bank