University of Victoria
Department of Economics

Econometrics Working Paper EWP0602

ISSN 1485-6441

ACCOUNTING FOR ACHIEVEMENT IN ATHENS: A COUNT DATA ANALYSIS OF NATIONAL OLYMPIC PERFORMANCE

Glen Roberts

Department of Economics, University of Victoria

Victoria, B.C., Canada V8W 2Y2

August, 2006

Abstract

We model summer Olympic medal counts using count data analysis. The advantage of this methodology is its explicit recognition of the discrete non-negative form of the dependent variable; *i.e.* the total number of medals won by a nation in a summer Olympiad. Using data from the most recent 2004 Summer Games in Athens, Poisson and negative binomial count data regression models are constructed. The chosen model is negative binomial and attaches statistical significance to Gross Domestic Product (GDP) *per capita*, the age dependency ratio, and a relatively cold winter climate. In contrast to previous studies, population, health expenditure *per capita*, and the effect of being a host or neighbour nation are all insignificant in explaining medal counts. We also find no "cricket effect" or "rugby effect."

Keywords: Olympic; count data; Poisson model; negative binomial model

JEL Classifications: C16; C25; L83

Author Contact:
Glen Roberts c/o David E. Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., V8W 2Y2, Canada; email: glenr@sfu.ca (cc: dgiles@uvic.ca); FAX: (250) 721-6214

I. Introduction

The International Olympic Committee (IOC), the governing body of the Olympic movement, proclaims the Olympic Games as a celebration of individual as opposed to national athletic achievement. Nevertheless, the celebration of Olympic success at the national level is inescapable, as epitomized by the ubiquitous presence of country medal standings in the media during any Olympiad. Naturally, the tremendous interest in national Olympic achievement has led researchers to explore the socioeconomic, cultural and geographic underpinnings of Olympic success. The intent of such research has not been to discount the primary importance of individual athletic talent, but rather to explore the fundamental factors affecting the ability of this talent to develop, flourish, and ultimately win medals at the Olympics.

This study proceeds in that spirit using count data econometric analysis: a methodology that, to the best of the author's knowledge, has never before been used to model Olympic performance. The unique advantage of this methodology is its explicit recognition of the discrete non-negative form of the dependent variable: the total number of medals won by a nation at a Summer Games. In this sense, count data analysis theoretically dominates other types of analysis which implicitly assume that the same dependent variable is continuous. Using data from the most recent 2004 Summer Games in Athens, Poisson and negative binomial (NB) count data regression models are constructed. The chosen model is negative binomial and attaches statistical significance to Gross Domestic Product (GDP) *per capita*, the age dependency ratio, and a relatively cold winter climate. In contrast to previous studies, population, health expenditure *per capita*, and the effect of being a host or neighbour nation are all insignificant in explaining medal counts. We also find no "cricket effect" or "rugby effect."

II. Background

Recent studies in this area have typically measured national Olympic success as some function of the medals won by a country either at a single Games or over multiple

Olympiads. Surprisingly, those studies in which the dependent variable has remained discrete have usually employed some form of analysis that assumes a continuous dependent variable. Hence, models arising from these studies have been inherently misspecified. Such studies include Hoffman *et al.* (2002), Johnson and Ali (2004) and Bian (2005), who all use Ordinary Least Squares (OLS) to explain total national medal counts.[1] Condon *et al.* (1999), who also use OLS, generate a discrete dependent variable by assigning an arbitrary number of points to the top eight placings in each competition.[2] Other studies, however, have avoided the use of a discrete dependent variable by employing some continuous measure of Olympic success. For example, Bernard and Busse (2004), by means of a Tobit analysis, examine determinants of total medal *shares*. Moosa and Smith (2004), who employ extreme bounds analysis, construct a continuous but arbitrarily weighted measure of the number of gold, silver, and bronze medals won.

These recent studies, despite varying methodologies, have also documented the explanatory power of numerous variables in determining Olympic success, including a country's total and *per capita* Gross Domestic Product (GDP), health expenditure *per capita*, population, political system, climate, and whether or not the country was a host or neighbour nation of the Games. For the most part, the potential effect of these variables on Olympic success is intuitively sensible. For example, we would expect that countries with relatively high living standards, as measured by GDP *per capita*, are likely to win a higher number of medals because they are more able to incur the costs of producing and supporting elite athletes.[3] A similar inference can be made regarding the effects of high health expenditure *per capita*. We would also have high medal count expectations of countries with bigger populations because, if nothing else, they have a larger base of potential medal-winning Olympians from which to choose. It also seems reasonable that

---

[1] This section summarizes only the dependent variables and econometric methodology used in the recent literature. We refrain from summarizing the *results* from *each* study because, by and large, the results are starkly similar in terms of the effects of variables on Olympic success. We discuss the typical effects of these variables below.

[2] Condon *et al.* also construct neural network models of Olympic success.

[3] Indeed, no known previous study has concluded otherwise. There has, however, been a debate over the appropriate exact measure: total GDP or GDP *per capita* (see Moosa and Smith, 2004). This paper will use GDP *per capita* because it is the more direct measure of living standards.

being a host or neighbour country of an Olympiad would provide various "home-court" advantages, including the ability to field a larger Olympic team due to reduced transportation costs.

The role of other variables is less obvious. Many previous studies have confirmed that countries with communist political systems have historically outperformed their counterparts, after controlling for other factors. However, in the post-Cold War era, only five communist states remain in existence. Furthermore, of these five, reliable comprehensive data exist for only one. As this paper examines medal counts from only the 2004 Summer Games, the effect of having a communist political system is no longer sensible or feasible to examine.[4] Another less obvious but potential explanatory variable is the climate of a country. All else being equal, it might be expected that colder countries would win fewer medals at the Summer Games. One argument is that Summer Olympiads consist of mostly outdoor events, and colder countries have less time to effectively train outdoors during the year. Furthermore, athletes from colder countries may not be as well acclimatized to competing in the severe heat that is common in Summer Olympiads. Despite this seemingly logical conjecture, however, recent studies examining the climate of a country have found conflicting results.[5]

There also exist other variables that, to the best of the author's knowledge, have been neglected in previous studies. The age dependency ratio; *i.e.*, the ratio of dependents to the working-age population, is one such variable. A higher age dependency ratio implies, firstly, that a relatively high number of potential medal-winning Olympians may be supporting their dependents in lieu of pursuing their Olympic dreams. Secondly, a high age dependency ratio usually implies a significant strain on social security programs, which in turn implies that less government resources are available for athletics.

[4] As supporting evidence of this claim, Moosa and Smith (2004), using data from the 2000 Summer Games, find the effect of being a communist nation to be insignificant.

[5] Johnson and Ali (2004) find that medals are decreasing in "warmth," whereas Hoffman *et al.* (2002) find an "inverted U" relationship. These studies did, however, employ very different measures of "warmth."

Finally, this paper examines another potential influence that has not been considered in previous studies. This potential effect is that of a country's national or most popular sport/s, in terms of participation and spectatorship, *not* being a competed event at the Olympics. If, in a particular country, a certain sport is unrivalled in popularity but it is not an Olympic sport, then we might expect that country to have a relatively low interest in the Olympics. With a relatively low interest in the Olympics, it is less likely that significant resources will be invested in Olympic sports in that country, implying a lower medal count. Furthermore, the country's best athletes will be predisposed to playing the popular non-Olympic sport/s, also implying a lower medal count. One may ask, given the substantial coverage of Olympic sports, for what sports this issue is relevant. There are two standouts: cricket and rugby union – almost certainly the most popular non-Olympic sports in the world. Cricket is widely played and hugely popular throughout the British Commonwealth. In countries such as India, Pakistan, Sri Lanka, and many Caribbean nations, the popularity of cricket is so overwhelming that it reduces almost all other sports to fringe activities. This may help explain why India, with a population exceeding one billion, only won a single medal in Athens. Rugby union, too, has a large worldwide following and is extremely popular in many countries. This paper examines whether there exists a significant "cricket effect" and/or "rugby effect."

III. Data[6]

Official Olympic medal data, obtained from the IOC, are from the most recent Summer Games in 2004. All socioeconomic data are from the most recent year (2000) for which reliable data exist for all relevant variables. GDP *per capita* data are in real (1996), PPP-adjusted terms. These, and the population data, are from Heston, Summers, and Aten, 2002. Data on health expenditure and age dependency ratios are from the World Bank World Development Indicators. Winter frost prevalence statistics; *i.e.*, data proxying for the "coldness" of a country, are from Masters and MacMillan, 2001. Host and neighbour nations are manually identified by the author. In contrast to other studies that examine multiple Games as a time-series, this paper examines only one Games and, therefore, is

[6] All data are available from the author upon request.

not afforded the privilege of separating the host and neighbour effects.[7] As a result, the host and neighbour countries are been combined into one host/neighbour dummy variable that is probably better interpreted as a "geographic advantage" variable.

Capturing the popularity of cricket and rugby is not easy, but world rankings are used as a proxy in each case. For cricket, a dummy variable has been constructed that takes a value of one for official Test-playing nations – the top ten cricket nations in the world.[8] In all of these nations, cricket is very popular. Moreover, there are only a handful of nations outside the Test-playing ones in which cricket is even moderately popular. By the same rationale, a rugby dummy variable has been constructed that takes a value of one for nations officially ranked in the world's top ten.[9] Cricket and rugby data are from the official world bodies: the International Cricket Council and the International Rugby Board.

Overall, cross-sectional data have been obtained for 133 out of the 201 nations that participated in the Athens Games. Of the missing countries, only eight won at least one medal, and only one country (Cuba) won more than five medals (27).

IV. Econometric Methodology

All of our models are Poisson or negative binomial (NB) count data regression models, estimated by maximum likelihood.[10] The dependent variable M_i is the total number of medals won by nation i ($i = 1, 2, \ldots, 133$). The vector of explanatory variables for nation i is given by $\boldsymbol{x}_i$, and the parameter vector to be estimated is denoted by $\boldsymbol{\beta}$. For the Poisson

---

[7] Clearly, since there is only one host, it makes no econometric sense to include a dummy variable for which only one observation takes a non-zero value.

[8] These ten test-playing "nations" actually correspond to eighteen countries (for which data are available) because the Caribbean countries play as a unified team (the West Indies) on the world stage.

[9] As of April 20, 2006. This corresponds to seven countries for which data are available. It should also be noted that these top ten rugby nations, with one or two exceptions, have not fallen outside the top ten for at least the past two decades.

[10] For comprehensive details of such models, see Cameron and Trivedi (1998) or Winkelmann (2000).

regression model, the expected medal count is modeled as a deterministic function of the set of explanatory variables and the unknown parameter vector. Mathematically,

$$E[M_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}] = Var[M_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}] = \lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}),$$

where λ_i is the mean (and variance) of the Poisson distribution. As the Poisson distribution is defined over only non-negative integer values, it is well suited to modeling discrete non-negative variables such as Olympic medal counts. The main disadvantage of the Poisson model, however, is its restriction of equidispersion; *i.e.*, the conditional mean equals the conditional variance. If the data are overdispersed, the NB regression model provides a popular alternative to the Poisson model. The NB distribution is also defined over only non-negative integer values; however, the NB is a two-parameter distribution and thus provides a richer modeling framework. For the NB regression model,

$$E[M_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}] = \mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}),$$

and

$$Var[M_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}] = \mu_i(1 + \eta^2\mu_i),$$

where η^2 is a parameter measuring the extent to which the conditional variance exceeds the conditional mean.[11] As the NB distribution converges to the Poisson distribution when $\eta^2 \to 0$, the Poisson and NB models are nested. Hence, the equidispersion restriction can be tested by a likelihood ratio (LR) test of the form: H_o: $\eta^2 = 0$; H_a: $\eta^2 \neq 0$.

V. Estimation Results

The estimation results (Table 2) present a preferred Poisson model and a preferred NB model. In each case, the preferred model is obtained by eliminating statistically insignificant (at the 10% level) regressors. The preferred Poisson model includes four significant explanatory variables: GDP *per capita* (GDP_PC), population (POP), the age

[11] In our output, estimated with the QMS EViews software package, $\log(\eta^2)$ is estimated and is referred to as the "shape" parameter.

dependency ratio (AGEDEP), and a relatively cold winter climate (FROST5). Apart from population, the preferred NB model includes the same regressors. The preferred Poisson and NB models are non-nested, so we use Akaike's Information Criterion (AIC) and clearly select the preferred NB model in favour of the preferred Poisson model.[12]

Our preferred NB model provides some interesting implications. As expected, GDP *per capita* has a statistically significant positive impact on medal counts; however, the mean marginal effect is hardly *practically* significant.[13,14] Also as expected, the age-dependency ratio has a statistically and practically significant negative impact on medal counts. This is a new result in the literature.

The preferred NB model also has some unexpected findings. Notably, population is insignificant in explaining national medal counts. This result conflicts with all known previous studies. One possible explanation, as indicated in other studies, is that there is a nonlinear relationship between population and medal counts. Specifically, it appears that there are highly diminishing returns to population, especially when we consider that participation quotas for each nation are imposed by the IOC.[15] The quotas generally result in smaller countries sending a higher proportion of their population to an Olympiad than larger countries. This at least somewhat diminishes the inherent advantage of a large population.

---

[12] In we omit POP from the preferred Poisson model in order to nest this model within the preferred NB model, we still reject the null of equidispersion; *i.e.*, we reject Ho: $\eta^2 = 0$. (p-val $\approx$ 0) The same result obtains if we *include* population in both models. (p-val $\approx$ 0) These results, if anything, further support the selection of the NB over the Poisson model.

[13] Marginal effects are computed at the sample means of the data as $\left[\frac{1}{n}\sum_{i=1}^{n} x_i'\hat{\beta}\right]\hat{\beta}_j$, where $\hat{\beta}_j$ is the estimated coefficient in question.

[14] The mean marginal effect of 2.347E-04 can be loosely interpreted as follows: an increase in *per capita* GDP (as measured in this study) of $10,000 is associated with an increase of approximately two Olympic medals.

[15] The decision process for participation quotas is kept confidential by the IOC.

Another surprising finding is that the mean marginal effect of FROST5 is significantly positive. This has the seemingly puzzling implication that colder countries perform better in medal counts at the Summer Olympic Games – the same result found by Johnson and Ali (2004). One possible explanation is that FROST5 is only a measure of coldness *in winter* (see Table 1). Very cold winters in themselves may not necessarily be detrimental to athlete training, so long as the summer months provide enough exposure to ideal outdoor training conditions. Athletes may also use the cold winter months for valuable rest and recuperation, as well as indoor training in state-of-the-art facilities.

Several other conclusions can be drawn from the preferred model. Health expenditure *per capita*, the host/neighbour nation effect, and the cricket and rugby effects all appear to hold no explanatory power. Although being a host or neighbour nation provides no explanatory power in this study, it is important to remember that we only examine data from one Olympiad. An examination of previous Olympiads could, as in other studies, easily conclude otherwise. While the cricket and rugby effects also hold no explanatory power in this study, we hope to inspire new lines of inquiry into (and alternative measures of) these effects.

VI. Conclusions

This paper uses count data analysis to explain national Summer Olympic medal counts. In contrast to most prior research on this topic, our use of count data analysis is theoretically pleasing because it explicitly recognizes the discrete non-negative form of the dependent variable. The preferred NB model indicates that GDP *per capita*, a relatively cold winter climate, and the age dependency ratio all have statistically significant impacts on national medal counts. Population, health expenditure *per capita*, the host/neighbour nation effect, and the cricket and rugby effects all appear to hold no explanatory power.

One direction for further research is a possible reconciliation of the insignificance of population in this paper's results. Another (possibly related) topic for further research would account for the large number of zero medal counts (69 in the sample), perhaps by employing a zero-inflated count data model.

Appendix

Table 1. ***Variable Summary***

Variable Abbreviation	**Variable Name**	**Comments**
TOTMED	Total Medals Won per Country	2004 Summer Olympic Games.
GDP_PC	Gross Domestic Product *per Capita*	PPP-adjusted (in real terms), 1996 international prices.
POP	Population	000s.
HEXP	Health Expenditure *per Capita*	Sum of total public and private expenditure on health care divided by the total population (current $U.S.).
AGEDEP	Age Dependency Ratio	Number of dependents (people younger than 15 or older than 64) to the working-age population (people ages 15-64).
FROST5	Frost Prevalence	Proportion of land with more than 5 frost-days per month in winter.
HOMENEIGHBOUR	Host or Neighbour Country	= 1 if host or neighbour country, 0 if not.
CRICKET	Test-Playing Cricket Nation	= 1 if Test-playing cricket country, 0 if not.
RUGBY	Top 10 Ranked Rugby Nation	= 1 if top 10 ranked rugby nation, 0 if not.

Table 2. ***Poisson and Negative Binomial (NB) Regression Results***

Variable	Regression with all Variables		Preferred Models			
	Poisson Coefficient (Std. Error)	NB Coefficient (Std. Error)	Poisson Coefficient (Std. Error)	Marginal Effect	NB Coefficient (Std. Error)	Marginal Effect
C	6.467 ** (1.678)	3.271 (1.803)	3.530 * (1.412)		3.134 ** (1.189)	
GDP_PC	-4.723E-05 (3.530E-05)	-3.558E-05 (6.069E-05)	5.271E-05 ** (2.035E-05)	3.483E-04	3.839E-05 * (1.961E-05)	2.347E-04
POP	1.792E-06 ** (2.877E-07)	4.471E-06 (5.859E-06)	2.146E-06 ** (3.207E-07)	1.418E-05		
HEXP	7.509E-04 ** (2.907E-04)	5.752E-04 (5.455E-04)				
AGEDEP	-9.786 ** (2.813)	-4.855 * (2.164)	-6.027 ** (2.141)	-39.836	-4.631 ** (1.769)	-28.307
FROST5	0.587 (0.403)	1.093 * (0.464)	1.051 * (0.499)	6.944	1.125 * (0.464)	6.873
HOMENEIGHBOUR	0.334 (0.414)	0.404 (0.391)				
CRICKET	-0.161 (0.355)	-0.262 (0.865)				
RUGBY	1.111 ** (0.391)	1.439 (0.776)				
"Shape" [a]		0.626 * (0.276)			0.897 ** (0.224)	
Log likelihood	-553.779	-279.109	-613.400		-288.100	
AIC [b]			1236.800		586.200	
Pseudo-R^2	0.590	0.793	0.546		0.787	

Note: Huber/White heteroskedastic-consistent standard errors are reported in parentheses.

* Statistically significant at the 5% level using two-tailed tests.

** Statistically significant at the 1% level using two-tailed tests.

a "Shape" = $\log[\eta^2]$.

b Akaike's Information Criterion (AIC) = -2log(*likelihood*) + 2(*number of parameters*).

Acknowledgements

I am very grateful to David Giles for his extensive advice on this paper.

References

Bernard, A. and Busse, M. R. (2004) Who Wins the Olympic Games: Economic Resources and Medal Totals, *Review of Economics and Statistics*, **86**, 413 - 417.

Bian, X. (2005) Predicting Olympic Medal Counts: The Effects of Economic Development on Olympic Performance, *The Park Place Economist*, **XIII**, 37 - 44.

Cameron, A. C. and Trivedi, P. K. (1998) *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

Condon, E. M., Golden, B. L., and Wasil, E. A. (1999) Predicting the success of nations at the Summer Olympics using neural networks, *Computers & Operations Research*, **26**, 1243 – 1265.

EViews (2004) *EViews 5 User's Guide*, Quantitative Micro Software, Irvine, CA.

Heston, A., Summers, R. and Aten, B. (2002) Penn World Table version 6.1, Center for International Comparisons, University of Pennsylvania (CICUP).

Hoffman, R., Ging, L. C. and Ramasamy, B (2002) Public policy and olympic success, *Applied Economics Letters*, **9**, 545 - 548.

International Cricket Council (2006) LG ICC Test Championship Rankings, http://www.icc-cricket.com/icc/test/.

International Olympic Committee (2006) ATHENS 2004, Games of the XXVIII Olympiad, http://www.olympic.org/uk/games/past/index_uk.asp?OLGT=1&OLGY=2004.

International Rugby Board (2006) Official IRB World Rankings, http://www.irb.com/WR/.

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

Masters, W. A., and McMillan, M. S. (2001) Climate and Scale in Economic Growth, *Journal of Economic Growth*, **6**, 167 - 186.

Moosa, I. A. and Smith, L. (2004) Economic Development Indicators as Determinants of Medal Winning at the Sydney Olympics: An Extreme Bounds Analysis, *Australian Economic Papers*, **43**, 288 - 301.

Winkelmann, R. (2000) *Econometric Analysis of Count Data* (Third Ed.), Springer, New York.

World Bank Group (2006) *World Development Indicators* (Online).