THE EFFECT OF PUBLIC SPORTING EXPENDITURES ON MEDAL SHARE AT THE SUMMER OLYMPIC GAMES:

A Study of the Differential Impact by Sport and Gender

ABSTRACT

This paper studies the effect of recreational and sporting services public expenditures on the medal shares of countries at the Summer Olympic Games. Using data on Olympic medals from 1996 to 2012 and a dynamic fixed effects Tobit model, this paper shows that the effect of public sporting expenditures on medal share is positive and significant for sports that generate lower revenues throughout the calendar year, and particularly for male athletes from lower income countries. Additionally, either currently hosting the Summer Olympic Games, or a country knowing it will host them four years later, positively impacts the medal share of female and male athletes of lower revenue-generating sports.

By: Joana Peixoto Nunes
Dr. Emily Marshall, Primary Supervisor
Dr. Anthony Underwood, Secondary Supervisor

Submitted in partial fulfillment of the Honors Requirements for the Economics Department

Dickinson College
Carlisle, Pennsylvania
May 2017

Table of Contents

1. Introduction

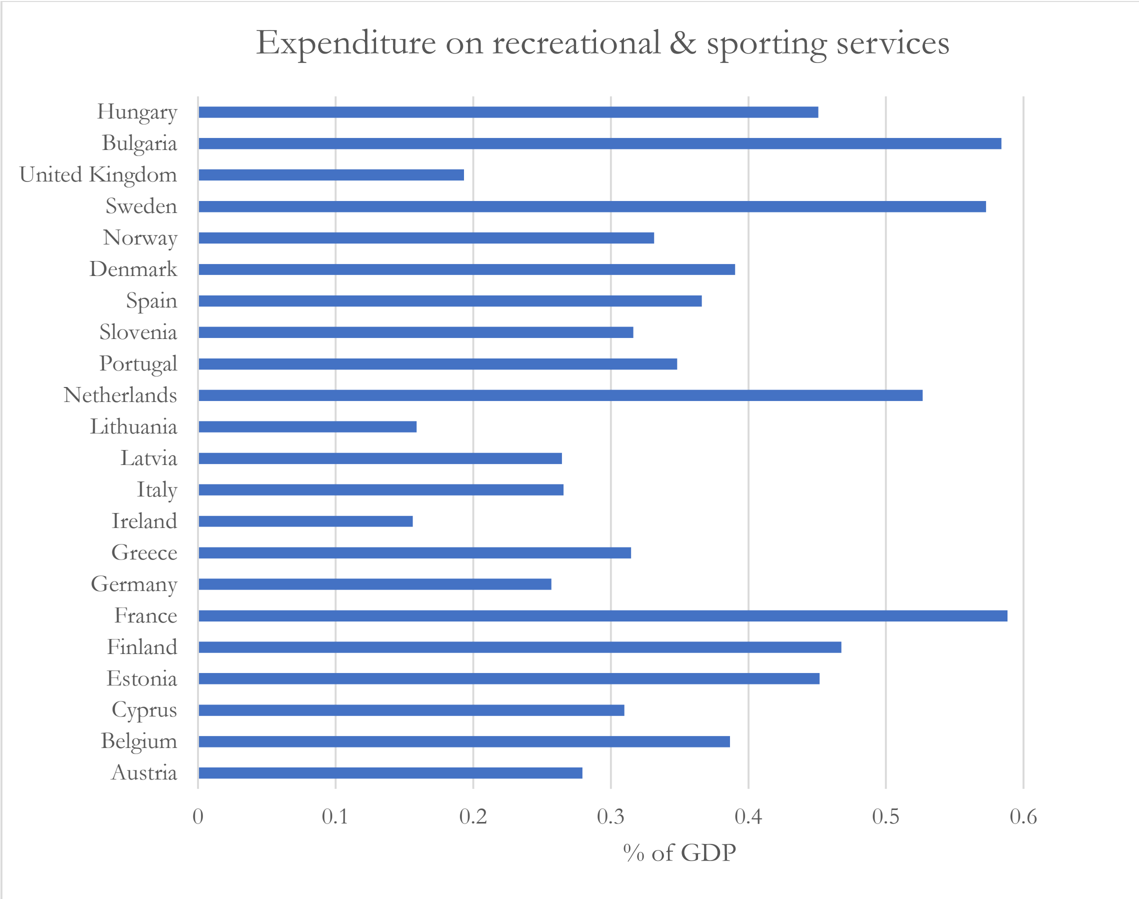
This paper estimates the impact that public spending on sports and sporting services has on the share of medals that countries win at the Summer Olympics. In addition, this paper examines whether this impact varies by type of sporting activity and gender of the sport. The government has a significant amount of choice in terms of expenditure allocation, so it can choose the degree to which it invests in sporting goods and services. Therefore, the findings of this paper could influence public policy and discretionary budget spending. Furthermore, the Summer Olympics (alongside the World Cup) is the largest worldwide sporting event in terms of interest and media coverage and there is constant demand for Olympic success forecasting (BBC Sport 2014). Moreover, a model that predicts medal wins for countries would also be beneficial to national teams for attracting sponsorships. Finally, the Summer Olympic Games are still not gender neutral – at the 2016 Rio Olympic Games there were 169 events for men and only 137 for women, so men naturally account for a higher medal share than women. In addition, the ratio of female to male participants is still below 0.5 (44% of athletes were women at the Rio 2016 Olympics) (WSJ 2016). So, comparing the impact of spending on men and women separately may reveal one of two things: either that for each $1 (USD) in spending on sports, the impact on female success is greater than that for men given the historical underrepresentation of women in the Games, or that the impact is greater for men, which would suggest that more than 50% of that $1 (USD) is being directed toward male athletes.

The contribution of this paper to the literature is the comparison across two different types of sports (according to their revenue-generating ability) and across gender. For that purpose, this paper analyzes four different groups of data. The main hypothesis is that higher public spending on recreation and sporting services increases the medal share of countries at the Summer Olympic Games. The additional hypothesis, not tested before in the literature, is that government spending on sporting-related activities will have a more significant impact on medal share won for sports that have a lower revenue-generating ability. The rationale is that for sports such as soccer and tennis—which are major spectator sports worldwide—there are numerous other events throughout the calendar year that attract many viewers and generate large revenues for their respective organization's teams, associations, and athletes. In contrast, the Summer Olympic Games are considered perhaps the main competition for sports such as swimming and rowing. Therefore, it is expected that sports federations of high revenue-generating sports do not need as much government funding and can thus create strong and competitive athletes regardless. The sports that are considered in this paper include the following: Swimming, Sailing, Shooting, Archery, Rowing, Wrestling, Hockey, Badminton, Fencing, Tennis, Athletics,[1] and Soccer.

Most countries spend less than 1% of their nominal GDP on sports and recreation annually, and for illustration purposes this paper shows a diagram from the IMF data bank. Note, also, that Olympic success is a zero sum game. If increasing public sporting expenditures increases a country's medal share, the medal shares of all other participating countries naturally decreases. Moreover, there may be a cap as to the magnitude of the funds that national governments are willing to devote to sports. Therefore, there may be a limit to the extent to which countries are willing to use public sporting expenditures as a means to increase their medal share beyond that of competing countries. Figure 1 displays expenditures on recreational and sporting services in 2015 as a percentage of GDP across a sample of different countries.

[1] Athletics includes track and field events, road running events, and race-walking events.

Figure 1: Expenditures, by country, on recreational and sporting services in 2015



It is interesting that the International Olympic Committee (IOC) generates over 40% of its Olympic revenues from commercial partnerships, precisely because it is a major international marketing platform. Other revenue sources include broadcast, ticketing, and licensing, with broadcast accounting for 47% of total Olympic Marketing Revenues (IOC 2017). This links to the idea that sports that attract more spectators will likely generate more revenues throughout the calendar year. In the specific case of the Summer Olympic Games, the IOC was predicted to generate a total of $9.3 billion in marketing revenues from the 2016 Rio Olympics only, which highlights the popularity of the Games (Independent 2016).

The remainder of the paper is organized as follows: Section 2 reviews the previous literature, Section 3 describes the data used and its organization, Section 4 describes the conceptual framework, the econometric model used, and the estimation methods, Section 5 presents the results, and Section 6 concludes the analysis. All auxiliary results are reported in the Appendix.

2. Literature review

The previous literature analyzes the variables that contribute to higher Olympic success. Using a Tobit model, Bernard and Busse (2004) evaluate the impact of population size and economic development on the number of medals countries won from 1960 to 1996, and conclude that GDP is the single most important determinant of medal success, regardless of population size. The authors also explore the role of government and political regimes in Olympic success and find that planned economies see a 1.7% increase in medal share[2] relative to unplanned economies. The same paper includes a production function for Olympic-caliber athletes within a country and uses a lagged medal share variable to represent Olympic talent as a durable but depreciating capital good – they capture depreciation by regressing medal share against the medal share won at the previous Games only. Additionally, they use the natural log of GDP per capita and the natural log of population size to form a regression model with medal share as their dependent variable. Bernard and Busse (2004) conclude that economic resources are important for generating quality athletes, but suggest that host countries generally win more medals than the number predicted by GDP alone. Using the Bernard and Busse (2004) framework, Forrest et al. (2010) further examine the effect of public spending on medal share. They include public spending on recreation as a covariate (rather than public spending on recreational and sporting services specifically as this paper does). Nonetheless, they find that the higher recreational spending of Australia, New Zealand, and The Netherlands improves their performance at the Olympic Games relative to what it would be if this paper

[2] Medal share is the number of medals won at the Olympic Games divided by the total number of medals awarded to all countries in all sports events in that same year. Medal share is calculated the same way throughout the literature, and the same method is applied in this paper.

took GDP alone as the explanatory factor. Forrest et al. (2010) also include a dummy for whether a country is hosting the games, and for whether it is a former soviet economy. The authors conclude that host and ex-Soviet bloc countries perform better than their GDP levels would have predicted. This paper also includes a dummy for whether a country is either currently hosting or whether it will host the following Games, but adds two new (although similar) covariates: the degree to which corruption is controlled for in the country, and the country's democratic index.

This paper adopts an extended version of the Bernard and Busse (2004) Cobb-Douglas production function for Olympic-caliber athletes that includes public spending on Recreational and Sporting Services in nominal USD terms $(Gov\ Exp\ Sports)$ as a "production" factor. This production function is used as a foundation for constructing a regression model, in which this paper includes the natural logs of public spending on Recreational and Sporting Services in nominal USD terms $(Gov\ Exp\ Sports)$ and population size $(Population)$ as regressors. Rather than including the actual GDP level, this paper uses a (high) income group $(High\ Income)$ dummy control variable instead and add a lagged medal share regressor $(Medal\ Share_{t-1})$. Additionally, and instead of the recurrent "Soviet sphere of influence" and "non-Soviet planned economy" dummies, this paper will introduce two new control variables that reflect each country's democratic and corruption control levels. This paper does so because many of the former Soviet economies no longer have centralized economies and are now democracies where sports success is not used as a propaganda strategy anymore.

Lui and Suen (2008) introduce country-specific effects and use a Poisson model (with medal count as the dependent variable) and a Tobit model with two new covariates (life expectancy and education). Education and life expectancy have no effect on Olympic success, but the coefficients on the lagged weighted sum of medals, population, and income, are all highly statistically and economically significant. This further supports the use of population, income, and lagged medal share as regressors.

Leeds and Leeds (2012) add to the literature by separating female and male performance in the 1996, 2000, 2004, and 2008 Olympics. Like Bernard and Busse (2004), Leeds and Leeds (2012) utilize the natural log of both population and real GDP per capita as regressors because of the diminishing marginal effects of higher population sizes and GDP levels. Leeds and Leeds (2012) assess how these variables influence medal winning differently by gender and find that higher fertility rates and later suffrage translate into fewer medals won by both women and men. In addition, the authors implement a negative binomial framework as opposed to a least squares regression or a Poisson model (to avoid the stringent equal mean and variance assumptions associated with the Poisson distribution – which the summary statistics in Table 2 will contradict). The most significant finding of Leeds and Leeds (2012) is that nations could potentially increase the medal share of both men and women by implementing policies that increase the political and economic participation of women.

This paper does not use variables that account for women's suffrage and fertility rates; however, it will implement a differential gender and sport type comparison by comparing four different models. The first model aggregates medal data for high revenue-generating Olympic events in which men participated, the second does the same for female Olympic events, the third considers low-revenue-generating sports in which men participated, and the fourth considers low-revenue generating sports in which women participated. If an equal share of funding towards sports were allocated between male and female athletes, then the hypothesis would be that public spending on sports has a more significant effect on Olympic success of women than on that of men. The rationale is similar to that of the cross-sport type comparison because the participation of women has increased over time as a result of political and economic changes.[3] These changes include the implementation of educational and training programs targeted at women, which are believed to be captured by public spending on sports data, though the most significant changes occurred prior to 1990 (i.e. outside the scope of this analysis) (IOC 2016). The similarity between the cross-sport type and cross-gender comparisons is the prediction that public spending on sports will likely have a greater effect on the traditionally "less supported" groups (i.e. women and sports that generate less revenues and spectators). However, and as mentioned previously, a higher proportion of public spending on sports likely goes to male athletes. Since this paper does not know where this spending is allocated, it may be that the impact of spending on sports is actually greater for male athletes than it is for women.

Hoffmann et al. (2002) consider the impact of annual average temperature and precipitation on Olympic success in order to assess whether living in "extreme" climates negatively affects outdoor sporting activity. Results indicate that the "optimal" average temperature is approximately 15 degrees Celsius.[4] The authors also consider per-capita GNP as a proxy for a country's sporting infrastructure and they also measure the willingness of a country to channel resources towards sports by measuring the current or recent existence of a socialist regime in the country. Their hypothesis is that socialist regimes are more likely to wish to see their country succeed at the Olympic Games and face fewer political obstacles in their investment decisions. Finally, Hoffmann et al. (2002) measure the degree to which a nation's culture values sports by counting the number of times a country has hosted the Olympic Summer Games between 1948 and 1996. The authors find whether the country has hosted the Olympics, population size, presence of a socialist regime, and GNP per capita to be statistically significant predictors of Olympic success. Hoffmann et al. (2002) conclude that medium-term policy strategies should include channeling governmental resources to sporting infrastructure

[3] Participation of women began in 1900 in tennis and golf events only, and expanded to other events through time (IOC, 2016).
[4] Examples of cities with 15 degrees Celsius as their average temperature include Rome, Melbourne, Atlanta, and Cape Town.

and improving a country's conditions for hosting the Olympics. The authors contribute to the literature by finding the optimal climatic conditions for athletes, conditions that are said to promote outdoor sporting activity. Drawing from the findings of Hoffmann et al. (2002), the current paper will also control for whether the country is currently hosting the Games or will host the following Games.

The proposed model will draw on the foundations provided by the literature and make the following additional contributions: the addition of a cross-sport type comparison (i.e. high vs. low revenue generating sports), a cross-gender comparison (i.e. male vs. female athletes), and the expansion of the Bernard and Busse (2004) Cobb-Douglas production function (by adding the public spending on recreational and sporting services as an additional "production" factor). Doing so allows us to observe the differential success across the four datasets previously mentioned based on the factors that influence medal share. Finally, this paper will include a corruption variable that ranks countries according to their relative control of corruption, a democracy index, and the Gini coefficient as additional explanatory variables in predicting a country's medal share. The idea is that countries with higher economic inequality likely have fewer Olympic caliber athletes than they otherwise would have if income were distributed more equally, since less well-off families could still afford to invest in their children's sporting activities. While previous literature has examined only the effect of being a former or present communist country, or of having a centrally planned economy, this paper will focus more on specific characteristics of the government. More corruption is hypothesized to lead to less public investment in sports infrastructure and public goods in general, and thus lead to lower sporting success. A more democratic country is hypothesized to be more accountable to the wishes of its people, and thus more likely to invest in public goods and services such as infrastructure for sporting activities, which is expected to yield a higher sporting success.

3. Data collection and organization

This section describes the dependent variable and independent variables used, as well as their descriptive statistics.

3.1 Data description

The data used in the present analysis consists of medal count data extracted from the International Olympics Committee (IOC) and from the Encyclopedia of the Games (EG). The variables used in the analysis are shown in Table 1.

Table 1: Variable Descriptions

Variable	Abbreviation	Description	Source
Medal share (M_t)	$Medal\ Share$	Observed medal share earned by each country in a given year (t) of the Summer Olympic Games. Medal share is the number of medals won at the Olympic Games divided by the total number of medals awarded to all countries in all sports events in that same year.	IOC and EG
Lagged medal share (M_{t-1})	$Medal\ Share_{t-1}$	Medal share earned by each country in the previous Summer Olympic Games	IOC and EG
The natural log of public spending on recreational and sporting services ($\ln(G)$)	$\ln(Gov\ Exp\ Sports)$	The natural log of total government expenditures on "Recreational and Sporting Services" in current U.S. dollars	IMF
The natural log of population size ($\ln(N)$)	$\ln(Population)$	The natural log of the total population size	World Bank
High income (HI)	$High\ Income$	Equal to 1 if the country is classified as high income (GNI per capita of $12,276 or more); equal to 0 otherwise	World Bank
Gini coefficient (GC)	$Gini$	Measure of a country's income distribution and inequality	World Bank
Corruption control (CC)	$Control\ of\ Corruption$	Control of Corruption – percentile rank A low rank means fewer (or less effective) measures are being taken to combat corruption relative to other countries	World Bank
Democracy index (DI)	$Democracy\ Index$	Voice and Accountability – percentile rank A low rank means the country is less democratic relative to other countries	World Bank
Olympic maturity (OM)	$Olympic\ Maturity$	Equal to 1 if the country has participated in any Summer Olympic Games before 1924; equal to 0 otherwise	IOC
Current or next host (CNH)	$Host$	Equal to 1 if the country is either hosting the current Summer Olympic Games or will host the next Summer Olympic Games; equal to 0 otherwise[5]	IOC

[5] Countries find out whether they will host the Olympics 7 years in advance. Therefore, in this year's Olympics, the host for the next Olympics is already known and is therefore likely to have channeled resources towards sports for the past 3 years already.

The independent variable of interest is "Expenditure on Recreational and Sporting Services"[6] in current U.S. dollars, a component of the more general public spending on "Recreation, cultural, and religious affairs."[7] Data pertaining to "Expenditure on Recreational and Sporting Services" ($Gov\ Exp\ Sports$) as a percentage of GDP is available for a panel of 91 countries from 1990-2015, but following Forrest et al. (2010), this paper uses this public spending on sports and sporting services spending variable in current U.S. dollars instead.[8] In addition, spending in actual USD amounts rather than as a percentage of GDP is likely more telling: a higher spending on sports as a percentage of GDP is not necessarily as significant as a lower spending on sports as a percentage of GDP if that GDP level is sufficiently higher than the former.

In order to implement the cross-sport comparison and determine if the effect of public spending on Olympic success is dependent on the type of sport and gender of the athletes, the Summer Olympic events are first divided into two separate groups: sports with revenues below the 25th and above the 75th percentiles. Based on 2013 global revenues for each international sport federation, a cumulative distribution was used to categorize sports by percentile.[9] The previous literature assumes that the effect of recreational expenditures is uniform across all

[6] This variable is intended to proxy for government spending on sports, as is standard in the literature (see Forrest et al. (2010)). The IMF defines "Recreational and Sporting Services," a subsection of "Recreation, culture, and religious affairs," as the "Provision of sporting and recreational services; administration of sporting and recreational affairs supervision and regulation of sporting facilities; operation or support of facilities for active sporting pursuits or events (playing fields, tennis courts, squash courts, running tracks, golf courses, boxing rings, skating rinks, gymnasia, etc.); operation or support of facilities for passive sporting pursuits or events (chiefly specially equipped venues for playing cards, board games, etc.); operation or support of facilities for recreational pursuits (parks, beaches, camping grounds and associated lodging places furnished on a non-commercial basis, swimming pools, public baths for washing, etc.); grants, loans or subsidies to support teams or individual competitors or players. Includes: facilities for spectator accommodation; national, regional or local team representation in sporting events. Excludes: zoological or botanical gardens, aquaria, arboreta and similar institutions (70820); sporting and recreational facilities associated with educational institutions (classified to the appropriate class of Division 709)."

[7] The IMF states that "Recreation, Culture, and Religion" includes "expenditures on services provided to individual persons and households and expenditures on services provided on a collective basis. Individual expenditures are allocated to groups 'Recreational and Sporting Services' and 'Cultural Services'; expenditures on collective services are assigned to groups 'Broadcasting and Publishing Services' and 'Recreation, Culture, and Religion'. Collective services are provided to the community as a whole. They include activities such as standards for providing recreational and cultural services; and applied research and experimental development into recreational, cultural and religious affairs and services."

[8] The IMF provides data on "Expenditure on Recreational and Sporting Services" and on "Recreation, cultural, and religious affairs" both as percentages of GDP. However, Forrest et al. (2010) convert recreation spending as a percentage of GDP into a monetary amount (by multiplying the variable by the GDP of each country in each year). This paper will do the same, and compute expenditure on sports as the product of the percentage of GDP allocated to recreation and sports and nominal GDP to arrive at a public spending variable in nominal USD terms.

[9] 2013 was chosen because it is a non-Olympic year, since revenues in Olympic years can be significantly higher for particular sports. The intention is to observe federations' generating abilities outside Olympic events.

sports. The current paper expands upon this by recognizing the potential for a differential impact based on the revenue generating ability of each sport, since lower revenue sports are hypothesized to be more dependent on government funding to create Olympic-level athletes Then, each of these two categories are further divided in two subgroups: men and women's sports. Hence, this paper will apply an econometric model to four different categories of athletes: male upper quartile sports, male lower quartile sports, female upper quartile sports, and female lower quartile sports. These four categories of athletes are likely not equally sensitive to government expenditures on sports, so this paper evaluates what these differences are and whether they are significant.

3.2 Summary statistics

This section shows the descriptive statistics of the four datasets: male upper quartile sports, male lower quartile sports, female upper quartile sports, and female lower quartile sports. The sports in the 25th percentile are Swimming, Sailing, Shooting, Archery, Rowing, Wrestling, Hockey, Badminton, and Fencing. Above the 75th percentile, the sports include Tennis, Athletics, and Soccer.

Table 2 presents the summary statistics for the dependent variable, $Medal\ Share$, by sport and gender. Note that each one of the four datasets (consisting of observations for the regressors identified in Table 1) is a panel of 23 countries over 5 years (1996, 2000, 2004, 2008, and 2012), since there is only one observation for the Summer Olympic Games every four years. The reason only results for 23 countries are displayed is that there is only data simultaneously for both $Gini$ and $\ln(Gov\ Exp\ Sports)$ for only 23 countries.[10]

[10] Although expenditure data for 91 countries is available, missing values for other covariates creates a sample of only 23 countries in each dataset. These 23 countries are listed in Appendix E.

Table 2: Summary Statistics (by sport and gender) of Medal Share

	Observations	Mean	Median	Std. Dev.	Min.	Max.
Upper Quartile - Female	65	0.3071%	0.0000%	0.0048	0.0000%	1.9380%
Upper Quartile -Male	65	0.3323%	0.0000%	0.0047	0.0000%	1.7921%
Lower Quartile - Female	65	1.6438%	0.3704%	0.0025	0.0000%	13.9706%
Lower Quartile - Male	65	1.5096%	0.6329%	0.0022	0.0000%	10.2990%
Number of Countries	23	23	23	23	23	23

The summary statistics in Table 2 show that no particular country within this paper's sample (in either their female or male Olympic events) earned a medal share superior to 13.97% of the total medals awarded in either the upper or lower quartile groups of sports. China won 13.97% of the total medals in the female lower quartile group in 2012 and won 10.30% of the medals awarded to the male lower quartile group in 2012. Note that $Medal\ Share$ is multiplied by 100 to arrive at the percentages shown in Table 2, and that the mean and median values for $Medal\ Share$ across all four groups are very low (mostly below 1%). This happens because there are many countries that do not win any medals in a specific year, but do in others so they are included in the model. In addition, the median is equal to zero for both upper quartile groups because at least half of the $Medal\ Share$ values in the two datasets are equal to zero.

Table 3 summarizes the descriptive statistics for the set of regressors considered and shows that the mean value for $\ln(Gov\ Exp\ Sports)$ is approximately 25, which implies that the average expenditure on recreational and sporting services across the sample of countries is approximately 72 billion USD in a given year. The values of the covariates (with the exception of $Medal\ Share_{t-1}$ which only differs from $Medal\ Share$ by one observation) are the same across all four datasets because the same countries and the same years are considered in all four groups.

Table 3: Summary statistics of all explanatory variables – common to all four datasets

Variables	Observations	Mean	Std. Dev.	Min.	Max.
$Gini$	65	29.4077	4.9118	21.6000	48.0000
$Control\ of\ Corruption$	65	84.6215	14.4465	38.8626	100.0000
$Democracy\ Index$	65	87.4351	14.1775	4.6948	100
$Olympic\ Maturity$	65	0.6308	0.4864	0.0000	1.0000
$Host$	65	0.0154	0.1240	0.0000	1.0000
$High\ Income$	65	0.8154	0.3910	0.0000	1.0000
$\ln(Gov\ Exp\ Sports)$	65	25.0965	1.8275	21.3920	29.1479
$\ln(Population)$	65	16.1703	1.3363	13.8897	21.0239

Due to the absence of observations for the recreation and sports expenditure variable before 1990, all observations before this date are dropped. Thus, each of the variables vary by country (i) and across time (t), with $t = 1996, 2000, 2004, 2008,$ and 2012. In addition, observations for which there are missing values are also dropped, which results in a short and balanced panel, with each group consisting of only 65 observations (since for some of the 91 observations for which there is $\ln(Gov\ Exp\ Sports)$ data, there is not always data pertaining to the other variables). This is unfortunate given that $\ln(Gov\ Exp\ Sports)$ is the variable of most interest to this paper's hypothesis. Since data for the variable of interest, $\ln(Gov\ Exp\ Sports)$, does not cover every single year and country, and since the Summer Olympic Games only take place every four years, more data pertaining to public sporting expenditures could significantly increase the set of observations, and thus benefit the analysis and conclusions presented here. Nonetheless, this paper can still verify the impact of government expenditure on recreation and sports across different sport groups and genders.

4. Conceptual framework, econometric models, and estimation methods

This section describes the econometric model and discusses the estimation methodology employed in this paper.

4.1. The regression model

Since success at the Summer Olympics depends on the number and quality of athletes in each individual country, this paper will build on the work of Bernard and Busse (2004) as a means of constructing the theory behind the regression model. Bernard and Busse (2004) use a Cobb-Douglas production function for Olympic-caliber athletes (T) in country i and year t as a function of the population(N), income(Y), and organizational ability (A):

$$T_{i,t} = A_{i,t} N_{i,t}^{\gamma} Y_{i,t}^{\theta}, \tag{1}$$

for $0 < \gamma, \theta < 1$.

For each country, the expected medal share must be a function of its Olympic-caliber athletes and hence, of its population size. However, the functional form of that relationship is unknown. Since the relationship between population size and medal share is non-linear (given that the impact of a greater population size diminishes as the population increases), Bernard and Busse (2004) use a log transformation to translate the production function into the determination of each country's expected medal share:[11]

$$E(Medal\ Share_{i,t}) = \ln \frac{T_{i,t}}{\sum_j T_{j,t}}, \tag{2}$$

where $Medal\ Share_{i,t}$ represents the medal share of country i in year t, and $\sum_j T_{j,t}$ is the sum of all Olympic athletes across all countries (for year t). The right-hand side of equation (2) measures the proportion of talented athletes in country i, at time t, relative to the number

[11] Countries do not send athletes in proportion to their population sizes because there are caps set by the International Olympics Committee (IOC).

of talented athletes across all countries at time t. Therefore, the assumption here is that country i's medal share is proportional to the fraction of Olympic athletes that that same country has, relative to the total number of Olympic athletes in all countries participating at the Summer Olympic Games at time t.

Instead of implementing equation (1), Bernard and Busse (2004) calculate the national Olympic-caliber talent as a linear combination of depreciated past talent and the investment in new talent (I). This paper follows a similar approach, but implements a log-log transformation such that:

$$\ln T_{i,t} = (1-\rho)\ln T_{i,t-1} + \ln I_{i,t}, \tag{3}$$

where $0 < \rho < 1$.

Again following Bernard and Busse (2004), a new production function is defined to represent investment in new talent (I); however, this paper replaces the income level with public spending on recreational and sporting services (G):

$$I_{i,t} = A_{i,t} N_{i,t}^{\gamma} G_{i,t}^{\delta}, \tag{4}$$

or

$$\ln(I_{i,t}) = \ln(A_{i,t}) + \gamma \ln(Population_{i,t}) + \delta \ln(Gov\ Exp\ Sports_{i,t}), \tag{5}$$

where $0 < \gamma, \delta < 1$. This paper replaces the income level variable because public spending on recreational and sporting services can likely explain investment in new talent more effectively. A country with a high income level does not necessarily invest in sports more than another country with a lower income level. If income is not being allocated to sports in a given country, then there is likely low correlation between income level and investment in new talent.

Furthermore, $\ln(A_i)$ is defined as a linear combination of the additional explanatory variables described in Table 1, i.e.:

$$\ln(A_{i,t}) = \beta_1 Gini_{i,t} + \beta_2 Control\ of\ Corruption_{i,t} + \beta_3 Democracy\ Index_{i,t} + \beta_4 Olympic\ Maturity_{i,t} + \beta_5 Host_{i,t} + \beta_6 High\ Income_{i,t}. \tag{6}$$

This paper includes all the variables shown in equation (6) in A because, like Bernard and Busse (2004), it is believed that there are other factors besides population size and income or spending variables that influence $Medal\ Share$ and that should not be left in the error term of the regression model. In their paper, Bernard and Busse (2004) include the following variables in their A term: one for whether the country is hosting the Games, two others to estimate the impact of being a planned economy, and a fourth one to estimate the impact of boycott years on medal success. For example, the U.S. boycotted the Summer Olympic Games in 1980, which benefited the Soviet Union, whereas the Soviet Union boycotted the games in 1984 to the advantage of the U.S.

The medal share function remains unchanged, given by equation (2). By combining equations (2), (3), (5), and (6), the regression model becomes:[12]

$$\begin{aligned} Medal\ Share_{i,t} &= \alpha + (1-\rho) Medal\ Share_{t-1} + \gamma \ln(Population_{i,t}) \\ &+ \delta \ln(Gov\ Exp\ Sports_{i,t}) + \beta_1 Gini_{i,t} \\ &+ \beta_2 Control\ of\ Corruption_{i,t} + \beta_3 Democracy\ Index_{i,t} \\ &+ \beta_4 Olympic\ Maturity_{i,t} + \beta_5 Host_{i,t} + \beta_6 High\ Income_{i,t} \\ &+ \varepsilon_{i,t}, \end{aligned} \tag{7}$$

where $\varepsilon_{i,t}$ is the error term for country i and year t, while α is the constant term of the regression model . Regression model (7) includes a lagged medal share variable because national Olympic talent is expected to be a function of both depreciated past talent and present talent.

[12] Please refer to Appendix F for more details. Refer also to Table 1 to associate variable names to their abbreviations.

4.2. The Tobit specification

The econometric model is a dynamic Tobit model. It is dynamic because the success of the previous Olympic Games is considered as a predictor of current success. Following the literature pertaining to the determinants of Olympic success, this paper implements a standard censored Tobit model (also known as a corner solution or a type 1 Tobit model). A type 1 Tobit model is applicable because the dependent variable, medal share, is continuous with a positive probability mass at zero (i.e. there are countries that do not win any medals at the Summer Olympic Games or that do not participate in some years). Much of the theory presented below is described in Wooldridge (2012). The type 1 Tobit model expresses the observed, dependent variable, $Medal\ Share$ (M), in terms of an underlying latent variable $Medal\ Share^*$ (M^*) such that:

$$\begin{aligned} Medal\ Share^*_{i,t} &= \alpha + (1-\rho) Medal\ Share_{i,t-1} + \delta \ln(Gov\ Exp\ Sports_{i,t}) \\ &+ \gamma \ln\left(Population_{i,t}\right) + \beta_1 Gini_{i,t} \\ &+ \beta_2 Control\ of\ Corruption_{i,t} + \beta_3 Democracy\ Index_{i,t} \\ &+ \beta_4 Olympic\ Maturity_{i,t} + \beta_5 Host_{i,t} + \beta_6 High\ Income_{i,t} \\ &+ \varepsilon_{i,t} \end{aligned} \tag{8}$$

where the error term associated with country i in year t is normally distributed with mean zero and variance σ^2. Equation (8) can be rewritten as:

$$Medal\ Share^*_{i,t} = \boldsymbol{x\beta} + \varepsilon_{i,t}, \tag{9}$$

where $E(Medal\ Share^*|\boldsymbol{x}) = \boldsymbol{x\beta}$ and $\boldsymbol{x}$ is a 1 x 10 vector of 9 explanatory variables, plus the intercept, while $\boldsymbol{\beta}$ is a 10 x 1 vector of the corresponding coefficients. Equation (9) absorbs the intercept for notational simplicity. Under these assumptions, the actual recorded medal share can be written as:

$$Medal\ Share_{i,t} = \max\left(0, Medal\ Share_{i,t}^{*}\right). \quad (10)$$

Since $Medal\ Share$ can only be greater or equal to 0, but $Medal\ Share^{*}$can assume negative values, the observed, true Medal Share is equal to the latent variable whenever the latent variable is greater or equal to 0. It is important to note that the variable that is observed is $Medal\ Share$, not $Medal\ Share^{*}$. $Medal\ Share^{*}$ is a latent variable (and an artificial construct) that satisfies the classical linear model assumptions. The coefficients obtained from the regression of $Medal\ Share^{*}$ describe the impact of $\boldsymbol{x}$ on $Medal\ Share^{*}$, not on $Medal\ Share$. The variable $Medal\ Share$ depends on $\boldsymbol{x}$ and $\boldsymbol{\beta}$, but in a nonlinear fashion. β_j measures the partial effects of $\boldsymbol{x_j}$ on $Medal\ Share^{*}$ (for $j = 1, \ldots, 9$). Equations (A.6) and (A.7) of Appendix A derive the marginal effects of any continuous independent variable x_j on $E(Medal\ Share|Medal\ Share > 0, \boldsymbol{x})$ and on $E(Medal\ Share|\boldsymbol{x})$, respectively. The difference between the two is that the former observes the marginal effects of the covariates strictly on positive values of $Medal\ Share$ (i.e. on countries that win at least one medal), whereas the latter considers the same effect but on all countries instead (regardless of whether they won at least one medal in a given year). These equations show that those marginal effects depend not only on β_j, but also on all other explanatory variables and parameters. If x_j is a binary variable, then $\frac{\partial E(Medal\ Share|x, Medal\ Share>0)}{\partial x_j}$ or $\frac{\partial E(Medal\ Share|x)}{\partial x_j}$ are computed for $x_j = 1$ and for $x_j = 0$, and the respective partial effects are given by the difference between the two results.

4.3. Estimation method

The Tobit model is estimated via Maximum Likelihood Estimation (MLE) due to the non-linearity in parameters. In this paper's model, a country will always have a minimum of 0 medals, but the use of a latent variable allows us to implement the Tobit model through MLE despite $Medal\ Share$ not following a normal distribution (see Figures 2 and 3).

Figure 2: Medal Share distribution – Lower Quartile

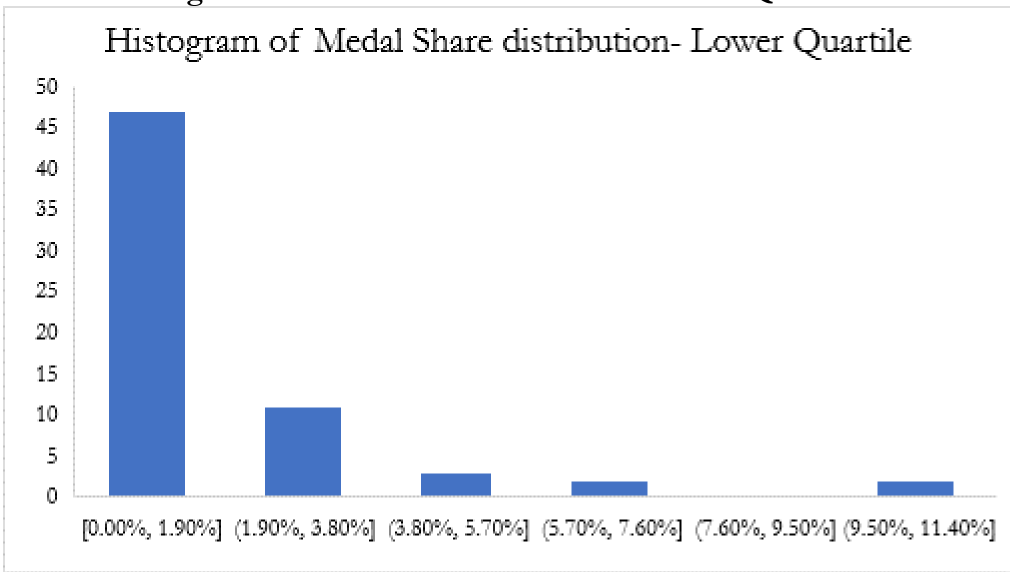
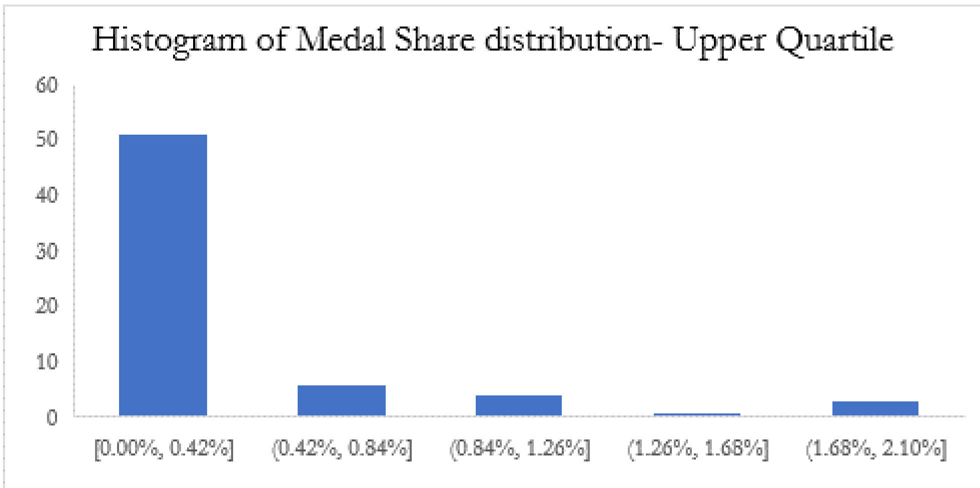


Figure 3: Medal Share distribution – Upper Quartile



Essentially, $M|\boldsymbol{x}$ does not follow a normal distribution because many of the observations are concentrated at zero, since many countries competing in the Olympics do not win any medals. As such, an Ordinary Least Squares (OLS) model is not appropriate. In addition, $E(M|\boldsymbol{x})$ is not linear in $\boldsymbol{x}$ unless the range of $\boldsymbol{M}$ is strictly limited. Finally, OLS implies constant marginal effects, which contradicts equations (A.6) and (A.7) of Appendix A that show that Tobit marginal effects also depend on $\boldsymbol{x}$. Therefore, the Tobit model must be estimated by a technique other than OLS.

MLE is an estimation strategy that can be used as an alternative to OLS. By writing the distribution of M as a function of the parameters to be estimated, MLE calculates the point estimates that maximize the likelihood of observing the sample at hand. Specifically, for this paper, MLE estimates the parameters $\boldsymbol{\beta}$ that maximize the joint probability densities $f\left(Medal\ Share_{i,t}\middle|\boldsymbol{\beta}\right)$ for all countries i and years t. Assuming that the observations are independent, this generalizes to the joint probability function given by equation (11):

$$\begin{aligned} &L(\boldsymbol{\beta}|Medal\ Share) \\ &= \prod_{i=1}^{n}\prod_{t=1}^{k} f\left(Medal\ Share_{i,t}\middle|\boldsymbol{\beta}\right) \\ &= f\left(Medal\ Share_{1,1}, Medal\ Share_{1,2}, \ldots, Medal\ Share_{n,k}\middle|\boldsymbol{\beta}\right), \end{aligned} \tag{11}$$

where "n" is the number of countries and "k" is the total number of years in the sample.

Since the vector $\boldsymbol{\beta}$ of parameters that maximizes L is the same $\boldsymbol{\beta}$ that maximizes the natural log of L, and since taking the natural log narrows the range of L, MLE maximizes the function $\ln\ L(\boldsymbol{\beta}|Medal\ Share)$ instead. The necessary condition for maximizing $\ln\ L(\boldsymbol{\beta}|Medal\ Share)$ is given by equating the first derivative of the natural log of equation (11) to zero:

$$\frac{\partial \ln\ L(\boldsymbol{\beta}|Medal\ Share)}{\partial \boldsymbol{\beta}} = \frac{\partial\left[\sum_{i=1}^{n}\sum_{t=1}^{k}\ln\left(f(Medal\ Share_{i,t}|\boldsymbol{\beta})\right)\right]}{\partial \boldsymbol{\beta}} = 0. \tag{12}$$

To estimate the Tobit model using MLE, the log-likelihood function for each observation (i,t) must be determined. For this purpose, note that when $Medal\ Share > 0$, $Medal\ Share = Medal\ Share^{*}$, which implies that $Medal\ share$ is distributed normally with mean $\boldsymbol{x\beta}$ and variance σ^2. Therefore, $P(Medal\ Share|Medal\ Share > 0, \boldsymbol{x}) =$

$\frac{1}{\sigma}\phi\left(\frac{Medal\ Share - \boldsymbol{x\beta}}{\sigma}\right)$, where ϕ denotes the probability density function of the standard normal distribution. Using equation (A.1) from Appendix A, the log-likelihood function for each observation (i, t) can then be written as:

$$l_{i,t}(\boldsymbol{\beta}, \sigma) = \ln\big(f(Medal\ Share_{i,t}|\boldsymbol{\beta})\big)$$

$$= \begin{cases} \ln\left(1 - \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\right), if\ Medal\ Share_{i,t} = 0 \\ \ln\left(\frac{1}{\sigma}\phi\left(\frac{M - \boldsymbol{x\beta}}{\sigma}\right)\right), if Medal\ Share_{i,t} > 0 \end{cases}, \quad (13)$$

where Φ represents the cumulative distribution function of the standard normal probability distribution. Finally, the log-likelihood for any sample size (n, k) can be obtained by summing $l_{i,t}(\boldsymbol{\beta}, \sigma)$ across all i and t.

4.4. Testing for the validity of the model specification

Before estimating the Tobit model, a simpler OLS formulation is implemented to gauge some intuition for the main effects of the model. As mentioned in section 4.1, the regression specification of interest is given by equation (7). The results of the OLS estimation are presented in Appendix B. Below is a discussion of key tests performed to arrive at the final regression specification.

Firstly, a simple OLS regression and variance inflation factor (VIF) test reveals that there are no significant collinearity issues among the independent variables. Next, the Hausman test is performed in order to determine if fixed effects or random effects are appropriate. Intuitively, it makes sense to presume that both panel-wide and time fixed effects are necessary. There are unobserved factors a_i that are unique to each country (e.g. cultural factors surrounding sports, the climate, etc.) that may affect the dependent variable ($Medal\ Share$) but are not controlled for unless a fixed effects model is implemented. In addition, time fixed effects should also be considered to control for situations such as global economic crises that affect all countries during a particular time period, especially since the data spans a lot of years. Tests for country and time fixed effects for each data set are reported in Table 4.

Table 4: Testing for country-level and time fixed effects

	Country fixed effects		Time fixed effects	
	χ^2 test stat.	P-value	F test stat.	P-value
Upper quartile sports - Male	30.8500	0.0001	0.2900	0.8832
Upper quartile sports - Female	18.0300	0.0210	0.9200	0.4660
Lower quartile sports - Female	117.9700	0.0000	0.3100	0.8678
Lower quartile sports - Male	38.7400	0.0000	0.4500	0.7698

In all four datasets, the Hausman test reveals that there is sufficient evidence to reject the null hypotheses that the difference in the coefficients under random and fixed effects is not systematic. Therefore, fixed effects at the country level are appropriate for all four groups. To test for time fixed effects, a joint significance test is used in which the null hypothesis is that all year level coefficients are jointly equal to zero. The p-values in all four datasets are large (greater than 0.4660), which implies that there is insufficient evidence to conclude that time fixed effects are necessary. Given these results, and the fact that the dependent variables are proportions ($Medal\ Share$), only country-level fixed effects are used. The fact that the dependent variable is not an absolute dollar amount suggests that time fixed effects are not as relevant as they would otherwise be; the dependent variable is constrained between 0 and 1, and thus, time-level effects cannot influence its maximum or minimum bounds. In addition, such time-level effects should not change the relative position of each country (in the medal share ranking), since those effects would be evenly spread amongst all countries. Therefore, the final model specification includes country-level fixed effects only.

Another assumption of the model is that there exists no correlation in the error terms across different time periods (e.g. the correlation between $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$is zero). Therefore, Table 5 tests for the presence of serial correlation.

Table 5: Testing for serial correlation (Wooldridge test)

	F test stat.	P-value
Upper quartile sports - Male	18.7200	0.0008
Upper quartile sports - Female	6.9960	0.0202
Lower quartile sports - Female	64.8680	0.0000
Lower quartile sports - Male	0.5980	0.4533

The null hypothesis of the serial correlation test in Table 5 is that the error terms are not serially correlated, and there is sufficient evidence to reject the null hypothesis that serial correlation is not present in three of the four groups. Serial correlation affects the efficiency of the OLS estimators; estimates of the standard errors will appear smaller than what they truly are and result in a tendency to reject the null hypothesis in instances where this should not be the case. However, these serial correlation tests should be interpreted with caution, since there is only a maximum of five time periods (1996, 2000, 2004, 2008, and 2012). So, this likely makes it more difficult to effectively capture a linear relationship between the error terms and their first-order lags, and hence neither the OLS or the Tobit models correct for serial correlation.

With panel data, it is important to note that even if the error variance does not change over time, it may change with the values of the regressors. Therefore, Table 6 reports the results of a Breusch-Pagan test that indicates that there is strong evidence to conclude that heteroscedasticity is present in all four regression models.

Table 6: Testing for the presence of heteroscedasticity

	Breusch-Pagan test	
	F test stat.	Critical value (5% significance level)
Upper quartile sports - Male	8.6521	3.1453
Upper quartile sports - Female	4.8230	3.1453
Lower quartile sports - Female	4.6035	3.1453
Lower quartile sports - Male	6.9716	3.1453

Therefore, standard errors that are robust to the presence of heteroscedasticity must be computed. The final regression specification uses robust standard errors. The OLS models for the four samples including country-level fixed effects and robust standard errors yield the results summarized in Table B.1 of the Appendix.

The OLS results are relegated to the appendix because they provide very limited information. A Tobit model is more appropriate, as it encompasses a latent variable that can assume both negative and positive values (which is not the case for the observed $Medal\ Share$ that only assumes nonnegative values with many observations at zero), which is important to meet the normality assumption. Hence, and as mentioned in the introduction, a corner solution Tobit model is more appropriate than an OLS regression, and the appropriateness of a Tobit specification is tested below by computing a Lagrange multiplier (LM) statistic. The alternative hypothesis of this test is that the model contains an error term that is heteroskedastic and non-normally distributed. Since the presence of both heteroscedasticity and non-normality implies that the Tobit model is inappropriate, rejecting the null hypothesis would suggest that a Tobit specification is not adequate. The test takes a Box-Cox transformation of the dependent variable, $\frac{Medal\ Share^{\lambda-1}}{\lambda}$, and tests the hypothesis that $\lambda = 1$ (Vncent, 2010). The LM-statistics are shown in Table 7.

Table 7: Testing the Tobit specification through an LM test

	LM test statistic	1% Bootstrap critical value	Is Tobit model suitable?
Upper quartile sports - Male	10.5580	14.4027	Yes
Upper quartile sports - Female	2.3605	18.9494	Yes
Lower quartile sports - Female	2.9828	17.4660	Yes
Lower quartile sports - Male	2.8141	14.7177	Yes

Comparing the test statistics contained in Table 7 to the 1% bootstrapped critical values, it is concluded that, for all four groups, there is not sufficient evidence to reject the hypothesis that the Tobit model is appropriate. Thus, a corner solution Tobit model is used as the final regression specification technique.

5. Results

In this section, results are reported from estimating the dynamic fixed effects Tobit model, and the marginal effects of interest are interpreted below. The first subsection considers the four datasets of interest: female athlete data for lower quartile sports, male athlete data for lower quartile sports, female athlete data for upper quartile sports, and male athlete data for upper quartile sports. The second subsection presents results for only two subgroups: one in which all male and female athletes competing in lower quartile sports are considered and another that combines all male and female athletes competing in upper quartile sports. The third subsection includes a new variable (Female Labor Force participation as a percentage of total females) as a robustness check.

5.1 Sports divided by percentile and gender

Before proceeding to the analysis of the Tobit model, the variables are tested for the presence of unit roots. Intuitively, it is reasonable to assume that there would be non-stationarity in the $\ln(Gov\ Exp\ Sports)$ and $\ln(Population)$ variables, which could create artificially high test statistics and suggests that there is a relationship between $\ln(Gov\ Exp\ Sports)$ and $Medal\ Share$, or between $\ln(Population)$ and $Medal\ Share$ when in fact there may not be one. Non-stationarity can be of three forms: trends, cycles, random walks, or a combination. In order for the Tobit results to be reliable, non-stationary variables must be transformed into stationary variables. A stationary process is one that is mean-reverting and has a constant variance over time. Non-stationary data with a deterministic trend occurs when one of the independent variables is related to a time variable, whereas a random walk with drift refers to a relationship between the variable's value at time t relative to its value at time $t-1$. Usually, non-stationarity can be removed by differencing in the presence of a drift, or subtracting the trend in the case of a non-stationary process with a deterministic trend (Hamilton, 1994).

In order to test for stationarity, an Augmented Dickey-Fuller test is conducted, which can be implemented in balanced panels and which runs an autoregressive model with a variable number of lags. For all the independent variables, the p-values were equal to, or almost equal to, 1.00 across all four groups. This implies that there is not enough evidence to reject the null, so the conclusion would be that unit roots are present in all of the independent variables. However, unit root tests are likely not applicable to this dataset. Although each dataset contains 65 observations, each panel (i.e. each country) only has a maximum of five time periods (and often even fewer in some cases due to missing values). Therefore, unit root tests are likely not very telling in this context.[13] The next part of the methodological discussion addresses whether random or fixed effects are appropriate in this context.

Implementing fixed effects with Tobit is not as straightforward as under OLS. Using fixed effects with Tobit increases the number of parameters to estimate (and, therefore,

[13] Given that the data only spans five Summer Olympic Games, one lag is used to test for the presence of unit roots. Any more lags would not allow the test to be performed, nor would it make intuitive sense to do so.

reduces the number of degrees of freedom); i.e. one additional parameter (a_i) for each country must be included in the log-likelihood function to be maximized. This issue does not occur with OLS because the time-constant effect a_i is eliminated through first differencing or time-demeaning. These two approaches do not work for the Tobit model because the relationship between the dependent and independent variables is nonlinear. Thus, the a_i's are not eliminated via fixed effects. However, country-level fixed effects are introduced by adding dummies for each country. This way, differences across countries in the observable or unobservable explanatory factors are controlled for and omitted variable bias is reduced. Fixed effects is used because utilizing random effects requires a very strong assumption—strict exogeneity and that the regressors are uncorrelated with any unobserved individual effects (e.g. climatic or demographic characteristics). In this context, that assumption is unreasonable. Therefore, the Tobit model with random effects is presented in Table C.1 of Appendix C and is not used for the primary analysis.

The most meaningful interpretation of the results comes from the marginal effects of the independent variables on $Medal\ Share$, rather than from the coefficients of the fixed effects Tobit regressions, since the latter highlight the effects on the latent variable $Medal\ Share^*$ instead of on the observed dependent variable. Table 8 below presents the coefficient estimates of the fixed effects Tobit model.

Table 8: Tobit Fixed Effects Results

Variables	Lower Quartile Female	Lower Quartile Male	Upper Quartile Female	Upper Quartile Male
$Medal\ Share_{t-1}$	-0.2541***	0.1188	-0.1482	-0.2653**
	(0.0811)	(0.1128)	(0.1115)	(0.1181)
$\ln(Gov\ Exp\ Sports)$	-0.0086***	0.0053*	-0.0028**	0.0005
	(0.0033)	(0.0031)	(0.0012)	(0.0011)
$\ln(Population)$	-0.0561	0.0086	0.0046	0.0000
	(0.0529)	(0.0526)	(0.0195)	(0.0173)
$High\ Income$	-0.0147*	-0.0014	0.0006	-0.0042
	(0.0084)	(0.0080)	(0.0031)	(0.0027)
$Gini$	-0.0014*	-0.0006	-0.0003	-0.0005*
	(0.0008)	(0.0008)	(0.0003)	(0.0003)
$Control\ of\ Corruption$	-0.0009	0.0009	-0.0003	-0.0002
	(0.0007)	(0.0006)	(0.0003)	(0.0002)
$Democracy\ Index$	0.0004	-0.0002	0.0002	-0.0002
	(0.0007)	(0.0007)	(0.0003)	(0.0002)
$Olympic\ Maturity$	0.1676	-0.0040	0.0131	0.0188
	(0.1054)	(0.1043)	(0.0390)	(0.0345)
$Host$	0.0165	0.0179*	-0.0025	-0.0071**
	(0.0102)	(0.0096)	(0.0039)	(0.0035)
Constant	1.2156	-0.3110	0.0118	0.0381
	(0.8144)	(0.8096)	(0.3014)	(0.2664)
Number of Countries	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

The coefficients presented in Table 8 represent the impacts of changes in the independent variables on the latent dependent variable, $Medal\ Share^*$, not on $Medal\ Share$. Therefore, in order to observe the impact of changes in the independent variables strictly on $Medal\ Share$, marginal effects are presented in Tables 9, 10, 11, and 12.

There are two marginal effects of interest: the marginal effects for the expected value of $Medal\ Share$ (i.e. the marginal effect on the actual variable of interest), and the marginal effects on positive observations of $Medal\ Share$. The former is represented as $\frac{\partial E(M|x)}{\partial x_j}$ and given by equation (A.7), whereas the latter is represented as $\frac{\partial E(M|x,M>0)}{\partial x_j}$ and follows from equation (A.6). The difference between the two lies in whether the sample is restricted to countries that won at least one medal in the Summer Olympics or in whether those that won zero medals in any given year are included as well. Note that the results of $\frac{\partial E(M|x)}{\partial x_j}$ will be equal to those that would be obtained in an OLS regression if the probability that an observation is different from zero is equal to one.

Table 9: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.1934***	0.0925	-0.1076	-0.2055**
	(0.0619)	(0.0878)	(0.0807)	(0.0911)
$\ln(Gov\ Exp\ Sports)$	-0.0066***	0.0041*	-0.0021**	0.0004
	(0.0025)	(0.0024)	(0.0009)	(0.0008)
$\ln(Population)$	-0.0427	0.0067	0.0033	0.0000
	(0.0402)	(0.0410)	(0.0142)	(0.0134)
$Gini$	-0.0011*	-0.0004	-0.0002	-0.0004*
	(0.0006)	(0.0006)	(0.0002)	(0.0002)
$Control\ of\ Corruption$	-0.0007	0.0007	-0.0002	-0.0001
	(0.0005)	(0.0005)	(0.0002)	(0.0002)
$Democracy\ Index$	0.0003	-0.0002	0.0002	-0.0001
	(0.0005)	(0.0005)	(0.0002)	(0.0002)
Country fixed effects	Yes	Yes	Yes	Yes
$High\ Income$	Yes	Yes	Yes	Yes
$Olympic\ Maturity$	Yes	Yes	Yes	Yes
$Host$	Yes	Yes	Yes	Yes
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

Table 9 indicates that in the male upper quartile group, only $Gini$ is statistically significant and negative, as expected. The marginal effect of $\ln(Gov\ Exp\ Sports)$ is positive, but insignificant. Nonetheless, for every 1% increase in $Gov\ Exp\ Sports$, $Medal\ Share$ increases by 0.0004%. The marginal effect is given as a percent change because a change in the variable $Medal\ Share$ is equivalent to a percentage point change in medals won. Given that the mean percentage of medals won in this group is 0.3323%, a potential increase of 0.0004% due to a 1% increase in $Gov\ Exp\ Sports$ is practically significant. The impact of public sporting expenditures on $Medal\ Share$ is positive among male athletes, and particularly lower quartile male sports. In fact, the impact of $\ln(Gov\ Exp\ Sports)$ for male athletes within lower quartile sports is both positive and significant, and larger relative to upper quartile sports. This difference is consistent with this paper's hypothesis: the effect is stronger within lower quartile sports because sports for which the Summer Olympic Games are the largest competition tend to be those that attract smaller crowds to competition, and thus likely generate lower revenues from ticketing, broadcasters, or sponsors. Hence, these lower quartile sports federations are likely to benefit more from government support relative to upper quartile sport federations. In addition, this positive impact of increases in $\ln(Gov\ Exp\ Sports)$ is only verified among male athletes. This shows that public money spent on sports is benefiting men's Olympic sporting success at the expense of increasing women's success too. Therefore, it is important for these funds to be equally distributed among female and male athletes so that participation and performance in female sports can increase.

The effect of public sporting expenditures on $Medal\ Share$ is negative for the female lower quartile group. Although the effect among female athletes of both lower and upper quartile sports is negative, the effect is more negative for lower quartile sports. The difference between how public sporting expenditures affects male and female sports could be highlighting the fact that there is no knowledge of the distribution of public spending on sports. If it is mostly directed to male athletes, then a percent increase in this spending has a positive effect on male athletes, as evidenced above. Similarly, the comparison across quartiles may be representing how these funds are supporting upper quartile sports at a disproportionate level relative to lower quartile sports.

The marginal effect of $\ln(Population)$ is positive across all groups except the female lower quartile. Nonetheless, it is not statistically significant in any group. Changes in the variables $Control\ of\ Corruption$ and $Democracy\ Index$ also fail to have a significant impact on Summer Olympic success. The coefficients on the $Medal\ Share_{t-1}$ are consistently negative (except for the lower quartile male group). The positive impact in the lower quartile male group is consistent with the model since Olympic talent is a durable good – i.e. athletes tend to compete for several continuous years. The marginal effects on strictly positive observations are shown in Table 10, and are very similar to those presented in Table 9.

Table 10: Tobit marginal effects $\left(\frac{\partial E(M|x,M>0)}{\partial x_j}\right)$

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.1684***	0.0795	-0.0884	-0.1716**
	(0.0537)	(0.0754)	(0.0665)	(0.0764)
$\ln(Gov\ Exp\ Sports)$	-0.0057***	0.0035*	-0.0017**	0.0003
	(0.0022)	(0.0021)	(0.0007)	(0.0007)
$\ln(Population)$	-0.0372	0.0057	0.0027	0.0000
	(0.0351)	(0.0352)	(0.0117)	(0.0112)
$Gini$	-0.0010*	-0.0004	-0.0002	-0.0003*
	(0.0006)	(0.0005)	(0.0002)	(0.0002)
$Control\ of\ Corruption$	-0.0006	0.0006	-0.0002	-0.0001
	(0.0004)	(0.0004)	(0.0002)	(0.0001)
$Democracy\ Index$	0.0003	-0.0001	0.0001	-0.0001
	(0.0005)	(0.0004)	(0.0002)	(0.0001)
Country fixed effects	Yes	Yes	Yes	Yes
$High\ Income$	Yes	Yes	Yes	Yes
$Olympic\ Maturity$	Yes	Yes	Yes	Yes
$Host$	Yes	Yes	Yes	Yes
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

Concerning dummy variables, more care must be taken in the interpretation, as it is not appropriate to find the marginal partial effects at the average. Instead, the marginal effects at both 0 and 1 are calculated, rather than at the average. Thus, this method is used to find partial effects of changes in the variables $High\ Income$, $Olympic\ Maturity$, and $Host$ as shown in Tables 11 and 12.

Table 11: Tobit marginal effects $\frac{\partial E(M|x)}{\partial x_j}$ for dummy variables

Variables		Value of x_j	Lower Quartile		Upper Quartile	
			Female	Male	Female	Male
$Host$		0	0.0178***	0.0161***	0.0038***	0.0039***
			(0.0008)	(0.0008)	(0.0003)	(0.0003)
		1	0.0327***	0.0327***	0.0022	0.0006
			(0.0099)	(0.0094)	(0.0019)	(0.0006)
	Difference		*0.0149*	*0.0166*	*-0.0016*	*-0.0033*
$High\ Income$		0	0.0287***	0.0174***	0.0034*	0.0068***
			(0.0067)	(0.0053)	(0.0018)	(0.0022)
		1	0.0163***	0.0162***	0.0038***	0.0033***
			(0.0011)	(0.0014)	(0.0005)	(0.0003)
	Difference		*-0.0124*	*-0.0012*	*0.0004*	*-0.0035*
Country fixed effects			Yes	Yes	Yes	Yes
Observations			65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

Table 12: Tobit marginal effects $\frac{\partial E(M|x,M>0)}{\partial x_j}$ for dummy variables

Variables	Value of x_j	Lower Quartile		Upper Quartile	
		Female	Male	Female	Male
$Host$	0	0.0194***	0.0178***	0.0045***	0.0044***
		(0.0008)	(0.0008)	(0.0003)	(0.0003)
	1	0.0329***	0.0327***	0.0032*	0.0016**
		(0.0095)	(0.0092)	(0.0017)	(0.0008)
Difference		*0.0135*	*0.0149*	*0.0029*	*-0.0028*
$High\ Income$	0	0.0292***	0.0187***	0.0042**	0.0070***
		(0.0063)	(0.0046)	(0.0015)	(0.0020)
	1	0.0181***	0.0177***	0.0045***	0.0039***
		(0.0010)	(0.0012)	(0.0004)	0.0003
Difference		*-0.0111*	*-0.0010*	*0.0003*	*-0.0031*
Country fixed effects		Yes	Yes	Yes	Yes
Observations		65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

The marginal effects of $Olympic\ Maturity$ are not estimable due to the lack of sufficient cases in which $Olympic\ Maturity = 1$. According to the marginal effects of $Host$ on all observations (Table 11), being the current or the next host of the Summer Olympic Games has a negative effect on $Medal\ Share$ across both male and female upper quartile sports. However, it has a positive and significant effect on lower quartile sports. This is consistent with the theory that if a particular country is hosting or will host the next Games, lower quartile sports are likely to benefit more from this as the hosting country will likely channel more funding into these sports to ensure that they perform their best for the home crowd. This same effect is visible in Table 12, when only strictly positive values of $Medal\ Share$ are considered.

The effect of being a high-income country has a positive and significant impact on $Medal\ Share$ only for the female upper quartile group. Although this variable is included for control purposes mainly, it may be indicating that high-income countries are more likely to support their female athletes and to support upper quartile sports compared to the reference group of low-income, lower middle-income, and upper middle-income economies. Anecdotally, this appears to be true for several countries. For example, soccer is one of the upper quartile sports, and the most successful teams are either from high-income or upper-middle income countries (e.g. Germany, Spain, Brazil, and Argentina). Among all observations for the female upper quartile group, the marginal effect of being a high-income country versus a low-income, lower middle-income, or a upper middle-income economy is a 0.0004% increase in $Medal\ Share$. Aside from being statistically significant, this appears to be a practically significant change as well.[14]

5.2 Sports divided by percentile only

In this subsection, the data is divided into only two groups: the lower quartile sports (for both male and female athletes) and the upper quartile sports (with both genders also). To observe potential gender and income effects, the regression model in equation (7) is expanded by adding three new variables: the proportion of medals awarded to women $(Female\ Medal\ Prop)$, permitting a differential effect of public sporting expenditures on female and male Olympic success; an interaction term between $Female\ Medal\ Prop$ and $\ln(Gov\ Exp\ Sports)$, and another interaction term between $High\ Income$ and $\ln(Gov\ Exp\ Sports)$. This final interaction term is introduced to evaluate the extent to which increasing levels of public sporting expenditures affects different economies in distinct manners.

[14] Several other model specifications were also considered. Beginning with a base model (in which only three main explanatory variables are included – $Medal\ Share_{t-1}$, $\ln(Gov\ Exp\ Sports)$, and $\ln(Population)$), other variables are then added incrementally. The tabulated results for each of the datasets are shown in Appendix D (Tables D.1, D.2, D.3, and D.4), where the marginal effects for the "best" model in each of the four datasets are included.

Using this new regression equation, a Tobit model with country fixed effects is estimated, and the marginal effects are shown in Tables 13 and 14.

Table 13: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$

Variables	Lower Quartile	Upper Quartile
$Medal\ Share_{t-1}$	0.0230	-0.3945***
	(0.0484)	(0.0992)
$\ln(Gov\ Exp\ Sports)$	0.0037*	-0.0011
	(0.0019)	(0.0025)
$\ln(Population)$	0.0722**	0.0363
	(0.0319)	(0.0411)
$Gini$	-0.0003	-0.0012**
	(0.0003)	(0.0005)
$Control\ of\ Corruption$	0.0001	-0.0010***
	(0.0003)	(0.0004)
$Democracy\ Index$	-0.0003	0.0004
	(0.0021)	(0.0004)
$Female\ Medal\ Prop$	0.0032	0.0102
	0.0021	0.0065
$Female\ Medal\ Prop \times \ln(Gov\ Exp\ Sports)$	6.21×10^{-15}	-7.09×10^{-14}***
	(7.03×10^{-15})	(-1.87×10^{-14})
$High\ Income \times \ln(Gov\ Exp\ Sports)$	-0.0097***	-0.0036
	(0.0023)	(0.0038)
Country fixed effects	Yes	Yes
$High\ Income$	Yes	Yes
$Olympic\ Maturity$	Yes	Yes
$Host$	Yes	Yes
Observations	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

Table 14: Tobit marginal effects $\left(\frac{\partial E(M|x,M>0)}{\partial x_j}\right)$

Variables	Lower Quartile	Upper Quartile
$Medal\ Share_{t-1}$	0.0205	-0.3446***
	(0.0430)	(0.0022)
$\ln(Gov\ Exp\ Sports)$	0.0033*	-0.0010
	(0.0017)	(0.0022)
$\ln(Population)$	0.0642**	0.0317
	(0.0283)	(0.0360)
$Gini$	-0.00031	-0.0010**
	(0.0283)	(0.0004)
$Control\ of\ Corruption$	0.0001	-0.0009***
	(0.0002)	(0.0003)
$Democracy\ Index$	-0.0003	0.0003
	(0.0003)	(0.0004)
$Female\ Medal\ Prop$	0.0029	0.0089
	0.0019	0.0057
$Female\ Medal\ Prop \times \ln(Gov\ Exp\ Sports)$	5.52×10^{-15}	-6.19×10^{-14}***
	(6.26×10^{-15})	(-1.63×10^{-14})
$High\ Income \times \ln(Gov\ Exp\ Sports)$	-0.0086***	-0.0031
	(0.0026)	0.0033
Country fixed effects	Yes	Yes
$High\ Income$	Yes	Yes
$Olympic\ Maturity$	Yes	Yes
$Host$	Yes	Yes
Observations	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

Consistent with the initial hypothesis, Tables 13 and 14 show that the marginal effect of $\ln(Gov\ Exp\ Sports)$ is now positive and statistically significant at the 10% level for the lower quartile sports, while negative (although statistically insignificant) for the set of upper quartile sports. Nevertheless, and given the presence of the two new interaction terms added, this interpretation is only appropriate if $Female\ Medal\ Prop = 0$ and $High\ Income = 0$ – i.e. the positive and statistically significant effect of $\ln(Gov\ Exp\ Sports)$ on $Medal\ Share$ applies to male athletes in lower income countries.

In fact, Table 15 shows that when $Female\ Medal\ Prop = 0$, a 1% increase in $Gov\ Exp\ Sports$ is expected to increase $Medal\ Share$ by 0.0037% in low income countries (i.e. when $High\ Income = 0$) and for lower quartile sports, which is economically very significant. In contrast, for high-income countries there is an expected decrease (-0.0060%). This is likely because high income countries already invest a significant amount in sports, whereas lower income countries may be allocating fewer resources to sports. Thus, a small increase in $\ln(Gov\ Exp\ Sports)$ might make a significant difference for lower income countries.

Table 15: Marginal effects $\left(\frac{\partial E(M)}{\partial \ln(Gov\ Exp\ Sports)}\right)$

	$Female\ Medal\ Prop = 0$		$Female\ Medal\ Prop = 1$	
	$High\ Income = 0$	$High\ Income = 1$	$High\ Income = 0$	$High\ Income = 1$
Lower Quartile	0.0037	-0.0060	0.0037	-0.0060
Upper Quartile	-0.0011	-0.0047	-0.0011	-0.0047

Table 16: Marginal effects $\left(\frac{\partial E(M)}{\partial \ln(Gov\ Exp\ Sports)}\right)$ at the mean

	$Female\ Medal\ Prop = mean$	
	$High\ Income = 0$	$High\ Income = 1$
Lower Quartile	0.0037	-0.0060
Upper Quartile	-0.0047	-0.0011

Note that the interaction between $Female\ Medal\ Prop$ and $\ln(Gov\ Exp\ Sports)$ yields relatively small coefficients, so the interpretation of the marginal effects when $Female\ Medal\ Prop$ is equal to either 0, 1, or its mean (please see Tables 15 and 16) are almost identical. Therefore, the more evident differences lie between high and low-income countries, and between lower and upper quartile sports. What these results show is that the impact of increases in public sporting spending on $Medal\ Share$ is positive for lower income countries and negative in high-income countries for lower quartile sports. In addition, there is a difference between upper and lower quartile sports: when $Female\ Medal\ Prop = 0$, the marginal effect of increases in spending is positive for lower quartile sports and negative for upper quartile sports within low-income countries. This may be an indication that it is these lower revenue-generating sports that need the most support, especially within lower income economies. Higher revenue-generating sports might do well without government funding given their popularity as sports and their ability to generate their own profits throughout the calendar year. Hence, lower-revenue sports athletes from low-income countries are more impacted at the margin by an additional percent increase in sports funding.

Overall, these results show that the only scenario (among the ones considered here) in which increases in $\ln(Gov\ Exp\ Sports)$ positively impacts $Medal\ Share$, occurs when lower income countries and lower quartile sports are considered. Therefore, government expenditures on sports could play a crucial role in enhancing the sporting success of lower quartile sports in lower income countries. Moreover, Tables 17 and 18 show that the impact of hosting the Olympic Games is positive and statistically significant within the lower quartile and negative otherwise. This may be because more funding is allocated to "less popular" events in which the hosting country has participating athletes. In the absence of hosting the Games, perhaps they would not have been as concerned about succeeding in those events.

Table 17: Tobit marginal effects $\frac{\partial E(M|x)}{\partial x_j}$ for dummy variables

Variables	Value of x_j	Lower Quartile	Upper Quartile
$Host$	0	0.0157***	0.0105***
		(0.0004)	(0.0006)
	1	0.0285***	0.0054
		(0.0053)	(0.0035)
Difference		*0.0128*	*-0.0051*
Country fixed effects		Yes	Yes
Observations		65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

Table 18: Tobit marginal effects $\frac{\partial E(M|x,M>0)}{\partial x_j}$ for dummy variables

Variables	Value of x_j	Lower Quartile	Upper Quartile
$Host$	0	0.0163***	0.0114***
		(0.0004)	(0.0006)
	1	0.0285***	0.0069**
		(0.0053)	(0.0032)
Difference		*0.0122*	*-0.0045*
Country fixed effects		Yes	Yes
Observations		65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

Finally, Table 19 shows the output of the joint significance tests. These are important given the presence of the two interaction terms. The results show that the marginal effect of $\ln(Gov\ Exp\ Sports)$ is statistically different from zero in both groups (upper and lower quartile sports), and that the marginal effect of $High\ Income$ is statistically different from zero within lower quartile sports—which supports the idea that less popular sports are likely not an investment priority of lower income countries.

Table 19: Joint significance tests

	χ^2 test stat.	p-value
$\frac{\partial \boldsymbol{E}(\boldsymbol{M})}{\partial \ln(Gov\ Exp\ Sports)} = \mathbf{0}$		
Upper Quartile	3.14	0.0766
Lower Quartile	8.44	0.0037
$\frac{\partial \boldsymbol{E}(\boldsymbol{M})}{\partial High\ Income} = \mathbf{0}$		
Upper Quartile	0.77	0.3810
Lower Quartile	10.44	0.0012
Number of Countries	23	23

5.3 Robustness Check

In this subsection, a new variable is added to the model: Female Labor Force participation (as a percentage of total females) – $Female\ LF\ Participation$. Because of the hypothesis that there is a gender bias in the allocation of public sporting expenditures, adding Female Labor Force participation as a control variable will serve as a robustness check. If the results are now different from what was shown in Subsections 5.1 and 5.2, it may be that the inclusion of this variable is controlling for some of the gender inequality present in the country. However, given that there is no gender inequality index specific to sports that covers the countries in this paper's sample, it may also be the case that Female Labor Force participation rates do not capture the gender bias present in public sporting expenditures.

In order to compare these new results from those of Subsections 5.1 and 5.2, Subsection 5.3 is further divided into two parts: first, the four dataset results are compared with the results of Subsection 5.1, and then the two dataset results are compared with the results of Subsection 5.2.

5.3.1 Sports divided by percentile and gender

When comparing Tables 9 and 20, it is evident that the results are very similar for the female lower quartile group. The public sporting expenditures variable, $\ln(Gov\ Exp\ Sports)$, is still both negative and significant at the 1% level, and all other explanatory variables retain the same sign. The only exception is the variable $Democracy\ Index$ that now has a negative sign (though it is statistically insignificant).

Table 20: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.2825***	-0.0693	-0.1039	-0.2021**
	(0.0626)	(0.0586)	(0.0829)	(0.0913)
$\ln(Gov\ Exp\ Sports)$	-0.0112***	0.0137***	-0.0021**	0.0002
	(0.0026)	(0.0034)	(0.0010)	(0.0009)
$\ln(Population)$	-0.1499***	0.1250**	0.0012	-0.0050
	(0.0479)	(0.0587)	(0.0178)	(0.0168)
$Gini$	-0.0016***	-0.0017**	-0.0002	-0.0004*
	(0.0006)	(0.0006)	(0.0002)	(0.0002)
$Control\ of\ Corruption$	-0.0002	0.0032***	-0.0002	-0.0001
	(0.0005)	(0.0005)	(0.0002)	(0.0002)
$Democracy\ Index$	0.0001	-0.0034***	0.0002	-0.0001
	(0.0005)	(0.0006)	(0.0002)	(0.0002)
$Female\ LF\ Participation$	0.0028***	-0.0007	0.0001	0.0001
	(0.0008)	(0.0006)	(0.0003)	(0.0003)
Country fixed effects	Yes	Yes	Yes	Yes
$High\ Income$	Yes	Yes	Yes	Yes
$Olympic\ Maturity$	Yes	Yes	Yes	Yes
$Host$	Yes	Yes	Yes	Yes
Observations	65	35	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

In the male lower quartile group, adding the variable $Female\ LF\ Participation$ does not change the sign of the $\ln(Gov\ Exp\ Sports)$ variable, but makes it statistically significant at the 1% level. The variable $Medal\ Share_{t-1}$ is no longer positive, but all other variables still have the same signs as in Subsection 5.1. However, it is important to note that there are only

35 observations in this male lower quartile group, due to the fact that there are missing values for the $Female\ LF\ Participation$ variable for the countries and years. Therefore, the greater statistical significance of the $\ln(Gov\ Exp\ Sports)$ variable should be interpreted with caution.

In the female upper quartile dataset, adding the variable $Female\ LF\ Participation$ does not change the interpretation of $\ln(Gov\ Exp\ Sports)$ as it is still negative and statistically significant at the 5% level, and all other marginal effects signs remain the same. Finally, in the male upper quartile group, it is also clear that the effect of changes in $\ln(Gov\ Exp\ Sports)$ remain positive and insignificant as in Subsection 5.1.

Therefore, adding the variable $Female\ LF\ Participation$ does not change our conclusions regarding the variable $\ln(Gov\ Exp\ Sports)$ as well as its impact across the two genders and across the two types of sports: it still has a positive impact on medal share won by men and a negative impact on that of women; and it still has a more positively significant impact on lower quartile male sports than on upper quartile male sports. This shows that the results in Subsection 5.1 are robust to the inclusion of the variable $Female\ LF\ Participation$. Consequently, it is likely that controlling for female labor force participation rates does not completely control for the gender bias present in the allocation of public sporting expenditures. In other words, there is likely a small correlation between women's participation in the labor force and the willingness of the government to invest in male and female athletes equally.

As shown in Table 21, the effect of a change in $Host$ from 0 to 1 is still negative for both upper quartile groups (male and female) and positive for the female lower quartile group. However, results for the male lower quartile group were not estimable due to the absence of sufficient observations at 1 for the $Host$ variable.

Table 21: Tobit marginal effects $\frac{\partial E(M|x)}{\partial x_j}$ for dummy variables

Variables		Value of x_j	Lower Quartile		Upper Quartile	
			Female	Male	Female	Male
$Host$		0	0.0176***		0.0038***	0.0039***
			(0.0008)		(0.0003)	(0.0003)
		1	0.0392***		0.0022	0.0007
			(0.0094)		(0.0020)	(0.0007)
	Difference		*0.0216*		*-0.0016*	*-0.0032*
				Not estimable		
$High\ Income$		0	0.0364***		0.0035*	0.0071***
			(0.0066)		(0.0019)	(0.0023)
		1	0.0155***		0.0038***	0.0033***
			(0.0008)		(0.0005)	(0.0003)
	Difference		*0.1190*		*0.0003*	*0.2929*
Country fixed effects			Yes		Yes	Yes
Observations			65		65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

The effect of a change in $High\ Income$ from 0 to 1 is still negative for both the female lower quartile and the male upper quartile groups, but is now slightly negative for the female upper quartile group. The results for the male lower quartile group were not estimable given the absence of sufficient observations at 0 for the $High\ Income$ variable.

5.3.2 Sports divided by percentile only

Table 22 shows that the lower quartile sports results now obtained are similar to those already shown in Subsection 5.2. The variable $\ln(Gov\ Exp\ Sports)$ is still positive and significant in the lower quartile group (though given the presence of interaction terms this effect is only relevant when $Female\ Medal\ Prop = 0$ and $High\ Income = 0$), and negative but insignificant in the upper quartile group (as in Subsection 5.2). In fact, all other variables keep the same sign and significance level, with the only exception being the marginal effect of changes in the $Control\ of\ Corruption$ variable that was negative and significant at the 1% level in the upper quartile group, but is now significant at the 5% level only. Nonetheless, the hypothesized sign of $Control\ of\ Corruption$ was a positive one instead.

Table 22: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$

Variables	Lower Quartile	Upper Quartile
$Medal\ Share_{t-1}$	0.0349	-0.3961***
	(0.0521)	(0.1019)
$\ln(Gov\ Exp\ Sports)$	0.0045*	-0.0010
	(0.0023)	(0.0029)
$\ln(Population)$	0.0876**	0.0383
	(0.0408)	(0.0503)
$Gini$	-0.0003	-0.0012**
	(0.0004)	(0.0005)
$Control\ of\ Corruption$	0.0002	-0.0010**
	(0.0003)	(0.0004)
$Democracy\ Index$	-0.0003	0.0004
	(0.0003)	(0.0004)
$Female\ Medal\ Prop$	0.0032	0.0101
	(0.0021)	(0.0066)
$Female\ Medal\ Prop\ \times \ln(Gov\ Exp\ Sports)$	7.58$\times 10^{-15}$	-7.09$\times 10^{-14}$***
	(7.37$\times 10^{-15}$)	(-1.87$\times 10^{-14}$)
$High\ Income \times \ln(Gov\ Exp\ Sports)$	-0.0103***	-0.0036
	(0.0031)	(0.0039)
$Female\ LF\ Participation$	-0.0003	-0.0000
	(0.0005)	(0.0006)
Country fixed effects	Yes	Yes
$High\ Income$	Yes	Yes
$Olympic\ Maturity$	Yes	Yes
$Host$	Yes	Yes
Observations	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at their mean value

To evaluate the impact of changes in $\ln(Gov\ Exp\ Sports)$ given the presence of interaction terms, Table 23 is presented. Comparing Table 23 to Table 15, it follows that the inclusion of $Female\ LF\ Participation$ does not change the signs of the marginal effects. There is still a positive impact of changes in $\ln(Gov\ Exp\ Sports)$ among lower quartile sports when $High\ Income = 0$ and when either $Female\ Medal\ Prop = 0$ or $Female\ Medal\ Prop =$

1 (given the small coefficient of the interaction term $Female\ Medal\ Prop \times \ln(Gov\ Exp\ Sports)$).

Table 23: Marginal effects $\left(\frac{\partial E(M)}{\partial \ln(Gov\ Exp\ Sports)}\right)$

	$Female\ Medal\ Prop = 0$		$Female\ Medal\ Prop = 1$	
	$High\ Income = 0$	$High\ Income = 1$	$High\ Income = 0$	$High\ Income = 1$
Lower Quartile	0.0035	-0.0068	0.0035	-0.0068
Upper Quartile	-0.0010	-0.0047	-0.0010	-0.0047

As shown in Table 24, the marginal effects of changes in $Host$ from 0 to 1 is still positive for the lower quartile group and negative for the upper quartile group, as was the case in section 5.2 without the variable $Female\ LF\ Participation$.

Table 24: Tobit marginal effects $\frac{\partial E(M|x)}{\partial x_j}$ for dummy variables

Variables	Value of x_j	Lower Quartile	Upper Quartile
$Host$	0	0.0157***	0.0105***
		(0.0004)	(0.0006)
	1	0.0275***	0.0053
		(0.0056)	(0.0035)
Difference		*0.0118*	*-0.0052*
Country fixed effects		Yes	Yes
Observations		65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All predictors at 0 and 1

Finally, the joint significance tests (shown in Table 25) show that the marginal effects of changes in $High\ Income$ and in $\ln(Gov\ Exp\ Sports)$ on $Medal\ Share$ are statistically different from zero in the same three groups as in Subsection 5.2.

Table 25: Joint significance tests

	χ^2 test stat.	p-value
$\frac{\boldsymbol{\partial E(M)}}{\boldsymbol{\partial} \ln(Gov\ Exp\ Sports)} = \mathbf{0}$		
Upper Quartile	3.02	0.0824
Lower Quartile	7.74	0.0054
$\frac{\boldsymbol{\partial E(M)}}{\boldsymbol{\partial} High\ Income} = \mathbf{0}$		
Upper Quartile	0.75	0.3861
Lower Quartile	10.60	0.0011
Number of Countries	23	23

In summary, and as was the case in Subsection 5.3.1, Subsection 5.3.2 confirms the robustness of the results of Subsection 5.2. The signs of the marginal effects remained the same, as did their statistical significance. Therefore, the gender bias present in the distribution of public sporting expenditures is not being captured by the new variable $Female\ LF\ Participation$.

6. Conclusion

This paper proposes an econometric model to predict success at the Summer Olympic Games that is based on economic theory and implemented through a dynamic fixed effects Tobit specification. This paper investigates the impact of recreational and sporting services government expenditures on a country's medal share for lower revenue-generating sports and for female athletes. The results show that female athletes are negatively impacted by public spending on sports, perhaps suggesting that much of this funding is channeled into activities that benefit male sports. However, given the absence of data that specifically outlines the distribution of public spending on sports, future research should aim to dissect the uses of public sporting expenditures.

Some additional differences across male and female and lower and upper quartile sports are noted. The marginal effects of the dummy explanatory variables are significant: there is a positive and significant effect of being a high-income country in the female upper quartile group, perhaps due to the greater availability of resources to support both male and female athletes. In addition, being a current host or a host of the following Summer Olympic Games has a positive and significant influence on lower quartile sports for both men and women. Finally, marginal effects for the male lower quartile group show a positive (though statistically insignificant) effect of public sporting expenditures. This supports the suspicion that male athletes may be receiving a disproportionate amount of government funds relative to female athletes. To control for gender inequality in the country as a whole (a form of it being in the allocation of public sporting funds), Subsection 5.3 contains a robustness check by including a proxy for gender inequality (i.e. $Female\ LF\ Participation$), but very similar results are obtained. Hence, the hypothesized gender bias in the allocation of public sporting expenditures might not be captured by the level of female participation in the labor force.

When combining both male and female athletes in each quartile, this paper shows that $Medal\ Share$ is positively impacted by increases in public sporting expenditures, particularly for lower income countries competing in lower quartile sports. The intuition is that lower income countries are more likely to be allocating fewer resources to sports to begin with, so a slight increase in spending on sports is more impactful. This is consistent with the hypothesis that public spending on sports is more important for sports federations that generate lower revenues throughout the year.

For future research on this topic, data on how public expenditures on sports are invested would help to explain the differential impact that is seen between female and male athlete Olympic success, and between medals earned in lower and upper quartile Olympic events. It is important that female and male athletes are supported equally and solely based on their talent and potential – not on their gender. This could be achieved through funding male and female athletes equally, as well as by increasing female empowerment, by promoting sports among females, or even by increasing the number of positions in sports occupied by women.

7. Appendix

The following appendices contain additional material on the partial effects of the Tobit model as well as additional econometric results.

7.1. Appendix A: Tobit partial effects

Given that the error term, $\varepsilon_{i,t}$, is normally distributed, then $\frac{\varepsilon}{\sigma}$ follows a standard normal distribution and, using equation (9), the following is obtained:

$$\begin{aligned} P(M=0|\boldsymbol{x}) &= P(M^{*}<0|\boldsymbol{x}) \\ &= P(\varepsilon < -\boldsymbol{x\beta}|\boldsymbol{x}) \\ &= P\left(\frac{\varepsilon}{\sigma} < \frac{-\boldsymbol{x\beta}}{\sigma}|\boldsymbol{x}\right) \\ &= \Phi\left(\frac{-\boldsymbol{x\beta}}{\boldsymbol{x}}\right) \\ &= 1-\Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right), \end{aligned} \tag{A.1}$$

where Φ symbolizes the cumulative distribution function of the standard normal. Moreover,

$$\begin{aligned} E(M|\boldsymbol{x}) &= P(M>0|\boldsymbol{x}) \times E(M|M>0,\boldsymbol{x}) + P(M=0|\boldsymbol{x}) \\ &\quad \times E(M|M=0,\boldsymbol{x}) \\ &= P(M>0|\boldsymbol{x}) \times E(M|M>0,\boldsymbol{x}) \\ &= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right) \times E(M|M>0,\boldsymbol{x}), \end{aligned} \tag{A.2}$$

where the last line follows from equation (A.1).

The following challenge is to expand $E(M|M>0,\boldsymbol{x})$. When $M>0$, $E(M|M>0,\boldsymbol{x}) = E(M^{*}|M^{*}>0,\boldsymbol{x})$, and that ε and $\boldsymbol{x}$ are independent**.** Therefore,

$$\begin{aligned}
\boldsymbol{E}(M|M>0,\boldsymbol{x}) &= \boldsymbol{E}(\boldsymbol{x\beta}+\varepsilon|M^{*}>0,\boldsymbol{x}) \\
&= \boldsymbol{x\beta}+E(\varepsilon|M^{*}>0) \\
&= \boldsymbol{x\beta}+E(\varepsilon|\varepsilon>-\boldsymbol{x\beta}) \\
&= \boldsymbol{x\beta}+\sigma E\left(\frac{\varepsilon}{\sigma}\middle|\frac{\varepsilon}{\sigma}>\frac{-\boldsymbol{x\beta}}{\sigma}\right).
\end{aligned} \tag{A.3}$$

If the probability density function of a standard normal variable z is denoted by $\phi(z)$, and if the fact that $\phi(-c)=\phi(c)$ and $E(z|z>c)=\frac{\phi(c)}{1-\Phi(c)}$, [15] is utilized, then equation (A.3) implies that

$$\begin{aligned}
E(M|M>0,\boldsymbol{x}) &= \boldsymbol{x\beta}+\sigma\left(\frac{\phi\left(\frac{-\boldsymbol{x\beta}}{\sigma}\right)}{1-\Phi\left(\frac{-\boldsymbol{x\beta}}{\sigma}\right)}\right) \\
&= \boldsymbol{x\beta}+\sigma\left(\frac{\phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}\right) \\
&= \boldsymbol{x\beta}+\sigma\lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right),
\end{aligned} \tag{A.4}$$

where $\lambda(c)=\frac{\phi(c)}{\Phi(c)}$ is known as the inverse Mills ratio (Wooldridge, 2012). The expected value that would be obtained from running an OLS model for observations where $M>0$ would merely be $\boldsymbol{x\beta}$, which contrasts with what is obtained in equation (A.4) – where the inverse Mills ratio is included. The inverse Mills ratio is considered an omitted variable that is usually correlated with $\boldsymbol{x}$ (Wooldridge, 2012). The Tobit and OLS estimates of the parameters would only be the same if $M>0$ for all (i,t).

If equations (A.2) and (A.4) are combined, the following is finally obtained:

[15] See for instance, Wooldridge (2012, p.598).

$$
\begin{aligned}
E(M|\boldsymbol{x}) &= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right) \times \left(\boldsymbol{x\beta} + \sigma\lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\right) \\
&= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\boldsymbol{x\beta} + \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\sigma\left(\frac{\phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}\right) \qquad \text{(A.5)} \\
&= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\boldsymbol{x\beta} + \sigma\phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right).
\end{aligned}
$$

In this analysis, both $E(M|\boldsymbol{x})$ and $E(M|M > 0, \boldsymbol{x})$ will likely be of interest, so the partial effects on both will be considered.

If x_j is a continuous variable, which is the case for some of the explanatory variables, then its partial effect on M can be determined as shown in equations (A.6) or (A.7) below. For this purpose, and using the fact that $\frac{\partial\lambda(c)}{\partial c} = -\lambda(c)[c + \lambda(c)]$ (Wooldridge, 2012, p.599), equation (A.4) implies that

$$
\begin{aligned}
\frac{\partial E(M|M > 0, \boldsymbol{x})}{\partial x_j} &= \beta_j + \sigma\left(\frac{\partial\lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\partial x_j}\right) \\
&= \beta_j + \sigma\left(\frac{\partial\lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\partial\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}\right) \times \left(\frac{\partial\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\partial x_j}\right) \qquad \text{(A.6)} \\
&= \beta_j + \sigma\left\{-\lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\left[\frac{\boldsymbol{x\beta}}{\sigma} + \lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\right]\right\}\left(\frac{\beta_j}{\sigma}\right) \\
&= \beta_j\left\{1 - \lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\left[\frac{\boldsymbol{x\beta}}{\sigma} + \lambda\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\right]\right\}.
\end{aligned}
$$

Following the same reasoning, equation (A.5) is differentiated:

$$\frac{\partial E(M|\boldsymbol{x})}{\partial x_j} = \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\beta_j + \frac{\partial \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\partial x_j}\boldsymbol{x\beta} + \sigma\frac{\partial \phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)}{\partial x_j}$$

$$= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\beta_j + \phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\frac{\beta_j}{\sigma}\boldsymbol{x\beta} - \sigma\phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\left(\frac{\boldsymbol{x\beta}}{\sigma} \times \frac{\beta_j}{\sigma}\right) \quad \text{(A.7)}$$

$$= \Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)\beta_j,$$

because $\frac{\partial \phi(f(z))}{\partial z} = -\phi(f(z))f(z)\frac{\partial f(z)}{\partial z}$ for any real valued function $f: \mathbb{R} \to \mathbb{R}$.[16] Likewise, the marginal effect of a change in an independent variable (x_j) on the probability that the $Medal\ Share$ is strictly positive (given all other explanatory variables) may also be of interest. Assuming that x_j is a continuous variable, and using equation (A.1), then

$$\frac{\partial P(M > 0|\boldsymbol{x})}{\partial x_j} = \frac{\beta_j}{\sigma}\phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right). \quad \text{(A.8)}$$

Equations (A.6) and (A.7) show that the partial effects of x_j on $E(M|M > 0, \boldsymbol{x})$ and on $E(M|\boldsymbol{x})$, respectively, not only depend on β_j, but also on all other explanatory variables and parameters. Note, too, that σ must be estimated.

There are two commonly used approaches for estimating these partial effects, given that the true values of the parameters are unknown. The two approaches are the partial effect at the average (PEA) and the average partial effect (APE) methods, with the latter being the preferred one. The APE approach estimates the scale factor $\Phi\left(\frac{\boldsymbol{x\beta}}{\sigma}\right)$ by computing $\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{\boldsymbol{x\hat{\beta}}}{\hat{\sigma}}\right)$. If x_j is discrete, both the PEA and APE methods may be implemented, with the APE approach still being the preferred one. If x_j is a binary variable, either

[16] This property follows because

$$\phi(f(z)) = \frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}f(z)^2\right],$$

and, therefore,

$$\frac{\partial \phi(f(z))}{\partial z} = \frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}f(z)^2\right]\left(-\frac{1}{2} \times 2f(z)\frac{\partial f(z)}{\partial z}\right).$$

$E(M|M > 0, \boldsymbol{x})$ or $E(M|\boldsymbol{x})$ is evaluated when $x_{j,t} = 1$ and when $x_{j,t} = 0$, and the respective partial effects are given by the differences between the two results.

7.2. Appendix B: OLS results

A simpler OLS formulation is implemented to gauge some intuition for the main effects of the regression model. Table B.1 summarizes the OLS regression results with country-level fixed effects and robust standard errors.

Table B.1: OLS regression results

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.2541	0.1188	-0.1482	-0.2653
	(0.1553)	(0.0913)	(0.1117)	(0.3677)
$\ln(Gov\ Exp\ Sports)$	-0.0086**	0.0053	-0.0028	0.0005
	(0.0043)	(0.0055)	(0.0021)	(0.0016)
$\ln(Population)$	-0.0561	0.0086	0.0046	0.0000
	(0.0662)	(0.0937)	(0.0321)	(0.0236)
$High\ Income$	-0.0147*	-0.0014	0.0006	-0.0042
	(0.0087)	(0.0093)	(0.0049)	(0.0035)
$Gini$	-0.0014	-0.0006	-0.0003	-0.0005
	(0.0016)	(0.0014)	(0.0007)	(0.0005)
$Control\ of\ Corruption$	-0.0009	0.0009	-0.0003	-0.0002
	(0.0011)	(0.0018)	(0.0006)	(0.0004)
$Democracy\ Index$	0.0004	-0.0002	0.0002	-0.0002
	(0.0010)	(0.0012)	(0.0005)	(0.0004)
$Olympic\ Maturity$	0.1676	-0.0040	0.0131	0.0188
	(0.1343)	(0.1966)	(0.0668)	(0.0464)
$Host$	0.0165***	0.0179***	-0.0025	-0.0071*
	(0.0035)	(0.0024)	(0.0017)	(0.0038)
Constant	1.2156	-0.3110	0.0118	0.0381
	(1.0280)	(1.6147)	(0.5080)	(0.3623)
Number of Countries	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes
Observations	65	65	65	65

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7.3. Appendix C: Tobit model with random effects

The following results differ from those of the Tobit model with fixed effects. This highlights the fact that there is heteroscedasticity present in this model, so the fixed effects model is robust to correlation between any country-level effects and the explanatory variables by removing any unobserved heterogeneity.

Table C.1: Tobit random effects results

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	0.5918***	0.1047	0.1152	0.260**
	(0.0884)	(0.2076)	(0.1204)	(0.1006)
$\ln(Gov\ Exp\ Sports)$	0.0049**	-0.0039	-0.0017**	0.0001
	(0.0022)	(0.0032)	(0.0008)	(0.0007)
$\ln(Population)$	-0.0009	0.0150**	0.0031***	0.0007
	(0.0027)	(0.0060)	(0.0011)	(0.0009)
$High\ Income$	-0.0060	-0.0019	0.0010	-0.0010
	(0.0056)	(0.0076)	(0.0021)	(0.0018)
$Gini$	-0.0002	-0.0001	0.0001	0.0001
	(0.0004)	(0.0005)	(0.0001)	(0.0001)
$Control\ of\ Corruption$	0.0002	0.0008**	0.0001	-0.0000
	(0.0002)	(0.0003)	(0.0001)	(0.0001)
$Democracy\ Index$	-0.0007***	-0.0010***	-0.0000	-0.0000
	(0.0002)	(0.0003)	(0.0001)	(0.0001)
$Olympic\ Maturity$	0.0049	-0.0031	0.0008	0.0024**
	(0.0033)	(0.0060)	(0.0013)	(0.0011)
$Host$	0.0125	0.0081	0.0014	0.0007
	(0.0113)	(0.0140)	(0.0046)	(0.0037)
Constant	-0.0491*	-0.1050**	-0.0178*	-0.0095
	(0.0291)	(0.0515)	(0.0107)	(0.0099)
Number of Countries	23	23	23	23
Country fixed effects	No	No	No	No
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7.4. Appendix D: Alternative model specifications

Aside from the full Tobit model with all explanatory variables, several other model specifications are also modeled. Beginning with a base case model (in which only three explanatory variables –$Medal\ Share_{t-1}$, $\ln(Gov\ Exp\ Sports)$, and $\ln(Population)$) are included, other variables are added incrementally. Tables D.1, D.2, D.3 and D.4 contain the results for the four sets of data.

Table D.1: Tobit results and comparisons amongst nested models for Male upper quartile sports

	Coefficient (standard error)				
Variables	Base Case	Model 2	Model 3	Model 4	Model 5
$Medal\ Share_{t-1}$	-0.1560	-0.1648	-0.1705	-0.2285*	-0.2285*
	(0.1173)	(0.1150)	(0.1167)	(0.1223)	(0.1223)
$\ln(Gov\ Exp\ Sports)$	-0.0006	0.0002	-0.0005	-0.0005	-0.0005
	(0.0009)	(0.0010)	(0.0010)	(0.0008)	(0.0008)
$\ln(Population)$	0.0182	0.0012	0.0175	0.0207	0.0207
	(0.0142)	(0.0179)	(0.0141)	(0.0140)	(0.0140)
$High\ Income$		-0.0032			
		(0.0028)			
$Gini$		-0.0005			
		(0.0003)			
$Control\ of\ Corruption$			-0.0001		
			(0.0002)		
$Democracy\ Index$			-0.00012		
			(0.0002)		
$Olympic\ Maturity$				-0.0256	
				(0.0270)	
$Host$				-0.0063*	-0.0063*
				(0.0037)	(0.0037)
Constant	-0.2740	-0.0098	-0.2443	-0.3162	-0.3162
	(0.2149)	(0.2719)	(0.2139)	(0.2116)	(0.2116)
Number of Countries	23	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Log Likelihood	288.8	290.1	289.5	290.2	290.2
Log-likelihood ratio test:		2.6900	1.4900	2.8700	2.8700
χ^2 test statistic					
(p-value)		0.2602	0.4744	0.09010	0.09010
Observations	65	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.2: Tobit results and comparisons amongst nested models for Female upper quartile sports

	Coefficient (standard error)				
Variables	Base Model	Model 2	Model 3	Model 4	Model 5
$Medal\ Share_{t-1}$	-0.1336	-0.1531	-0.1406	-0.1099	-0.1099
	(0.1044)	(0.1048)	(0.1059)	(0.1092)	(0.1092)
$\ln(Gov\ Exp\ Sports)$	-0.0024***	-0.0024**	-0.0028***	-0.0024**	-0.0024**
	(0.0009)	(0.0011)	(0.0011)	(0.0009)	(0.0009)
$\ln(Population)$	0.0064	0.0031	0.0076	0.0072	0.0072
	(0.0153)	(0.0194)	(0.0154)	(0.0153)	(0.0153)
$High\ Income$		0.0006			
		(0.0030)			
$Gini$		-0.0003			
		(0.0003)			
$Control\ of\ Corruption$			0.0002		
			(0.0002)		
$Democracy\ Index$			-0.0001		
			(0.0002)		
$Olympic\ Maturity$				0.0040	
				(0.0296)	
$Host$				-0.0028	-0.0028
				(0.0040)	(0.0040)
Constant	-0.0389	0.0176	-0.0468	-0.0520	-0.0520
	(0.2316)	(0.2957)	(0.2325)	(0.2314)	(0.2314)
Number of Countries	23	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Log Likelihood	284.0	284.7	284.2	284.2	284.2
Log-likelihood ratio test: χ^2 test statistic		1.45	0.42	0.51	0.51
(p-value)		0.4836	0.8123	0.4755	0.4755
Observations	65	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.3: Tobit results and comparisons amongst nested models for Male lower quartile sports

	Coefficient (standard error)				
Variables	Base Model	Model 2	Model 3	Model 4	Model 5
$Medal\ Share_{t-1}$	0.1256	0.1441	0.1112	0.1190	0.1190
	(0.1120)	(0.1138)	(0.1127)	(0.1099)	(0.1099)
$\ln(Gov\ Exp\ Sports)$	0.0035	0.0048	0.0049*	0.0033	0.0033
	(0.0025)	(0.0031)	(0.0027)	(0.0024)	(0.0024)
$\ln(Population)$	0.0323	0.0063	0.0294	0.0274	0.0274
	(0.0413)	(0.0540)	(0.0407)	(0.0406)	(0.0406)
$High\ Income$		-0.0033			
		(0.0081)			
$Gini$		-0.0008			
		(0.0008)			
$Control\ of\ Corruption$			0.0009		
			(0.0006)		
$Democracy\ Index$			-0.0003		
			(0.0007)		
$Olympic\ Maturity$				-0.0423	
				(0.0783)	
$Host$				0.0157	0.0157
				(0.0098)	(0.0098)
Constant	-0.6003	-0.1938	-0.6427	-0.5180	-0.5180
	(0.6241)	(0.8204)	(0.6160)	(0.6143)	(0.6143)
Number of Countries	23	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Log Likelihood	221.2	221.7	222.3	222.4	222.4
Log-likelihood ratio test: χ^2 test statistic		1.11	2.2	2.52	2.52
(p-value)		0.5743	0.3336	0.1121	0.1121
Observations	65	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.3 shows all models tested for the male lower quartile group provide coefficients with the expected coefficient sign for almost every regressor that was tested, namely for $Medal\ Share_{t-1}$, $\ln(Gov\ Exp\ Sports)$, for $\ln(Population)$, for $Gini$, for $Control\ of\ Corruption$, and for the $Host$ dummy variable. Unfortunately, such empirical evidence in support of the initial hypotheses is not statistically significant (except for the main explanatory variable, $\ln(Gov\ Exp\ Sports)$, which is significant at the 10% level under Model 3).

Table D.4: Tobit results and comparisons amongst nested models for Female lower quartile sports

	Coefficient (standard error)				
Variables	Base Model	Model 2	Model 3	Model 4	Model 5
$Medal\ Share_{t-1}$	-0.2299***	-0.2492***	-0.2227***	-0.2360***	-0.2360***
	(0.0841)	(0.0838)	(0.0836)	(0.0825)	(0.0825)
$\ln(Gov\ Exp\ Sports)$	-0.0101***	-0.0075**	-0.0116***	-0.0104***	-0.0104***
	(0.0026)	(0.0031)	(0.0030)	(0.0026)	(0.0026)
$\ln(Population)$	0.0091	-0.0422	0.0130	0.0036	0.0036
	(0.0434)	(0.0542)	(0.0431)	(0.0426)	(0.0426)
$High\ Income$		-0.0125			
		(0.0085)			
$Gini$		-0.0010			
		(0.0009)			
$Control\ of\ Corruption$			-0.0008		
			(0.0007)		
$Democracy\ Index$			0.0004		
			(0.0007)		
$Olympic\ Maturity$				0.0395	
				(0.0824)	
$Host$				0.0171	0.0171
				(0.0104)	(0.0104)
Constant	0.1160	0.9029	0.1258	0.2087	0.2087
	(0.6549)	(0.8245)	(0.6514)	(0.6443)	(0.6443)
Number of Countries	23	23	23	23	23
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Log Likelihood	216.9	218.0	217.6	218.2	218.2
Log-likelihood ratio test: χ^2 test statistic		2.32	1.5	2.63	2.63
(p-value)		0.5743	0.3336	0.1121	0.1121
Observations	65	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Finally, Tables D.5, D.6, D.7 and D.8 include the marginal effects for the "best" model in each of the four datasets. The "best" model used in each of the Tables D.1, D.2, D.3 and D.4 is chosen through likelihood ratio tests.

Table D.5: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.1754***	0.0925	-0.1076	-0.2055**
	-0.0643	(0.0878)	(0.0807)	(0.0911)
$\ln(Gov\ Exp\ Sports)$	-0.0077***	0.0041*	-0.0021**	0.0004
	(0.0020)	(0.0024)	(0.0009)	(0.0008)
$\ln(Population)$	0.0069	0.0067	0.0033	0.0000
	(0.0331)	(0.0410)	(0.0142)	(0.0134)
Country fixed effects	Yes	Yes	Yes	Yes
$Gini$	No	No	No	No
$Control\ of\ Corruption$	No	No	No	No
$Democracy\ Index$	No	No	No	No
$High\ Income$	No	No	No	No
$Olympic\ Maturity$	No	No	No	No
$Host$	No	No	No	Yes
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.6: Tobit marginal effects $\left(\frac{\partial E(M|x,M>0)}{\partial x_j}\right)$

	Lower Quartile		Upper Quartile	
Variables	Female	Male	Female	Male
$Medal\ Share_{t-1}$	-0.1513***	0.0832	-0.0791	-0.1457*
	(0.0553)	(0.0741)	(0.0617)	(0.0779)
$\ln(Gov\ Exp\ Sports)$	-0.0067***	0.0023	-0.0015**	-0.0003
	(0.0018)	(0.0016)	(0.0006)	(0.0005)
$\ln(Population)$	0.0060	0.0214	0.0038	0.0132
	(0.0285)	(0.0274)	(0.0091)	(0.0089)
Country fixed effects	Yes	Yes	Yes	Yes
$Gini$	No	No	No	No
$Control\ of\ Corruption$	No	No	No	No
$Democracy\ Index$	No	No	No	No
$High\ Income$	No	No	No	No
$Olympic\ Maturity$	No	No	No	No
$Host$	No	No	No	Yes
Observations	65	65	65	65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.7: Tobit marginal effects $\left(\frac{\partial E(M|x)}{\partial x_j}\right)$ for dummy variable

Variable (x_j)	Value of x_j	Upper Quartile Male
$Host$	0	0.0039
	1	0.0008
Difference		*-0.0031*
Country fixed effects		Yes
Observations		65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at 0 and 1

Table D.8: Tobit marginal effects $\left(\frac{\partial E(M|x,M>0)}{\partial x_j}\right)$ for dummy variable

Variable (x_j)	Value of x_j	Upper Quartile Male
$Host$	0	0.0045
	1	0.0018
Difference		*-0.0027*
Country fixed effects		Yes
Observations		65

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: All predictors at 0 and 1

7.5. Appendix E: Countries included in each dataset

Each dataset includes 23 countries. These include: Austria, Belgium, Bulgaria, China, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, and the United Kingdom.

7.6. Appendix F: Proof of equation (7)

Combining equations (2), (3) and (5), then:

$$\begin{aligned} E(M_{i,t}) &= \ln\left(T_{i,t}\right) - \ln\sum_j T_{j,t} \\ &= (1-\rho)\ln T_{i,t-1} + \ln\left(A_{i,t}\right) + \gamma\ln\left(N_{i,t}\right) + \delta\ln\left(G_{i,t}\right) \\ &\quad - \ln\sum_j T_{j,t}. \end{aligned} \tag{F.1}$$

Using equation (2) again, and replacing $\ln\left(A_{i,t}\right)$ by the right-hand side of equation (6), equation (F.1) is rewritten as

$$\begin{aligned} E(M_{i,t}) &= (1-\rho)\left(E(M_{i,t-1}) + \ln\sum_j T_{j,t-1}\right) + \beta_1 GC_{i,t} + \beta_2 CC_{i,t} \\ &\quad + \beta_3 DI_{i,t} + \beta_4 OM_{i,t} + \beta_5 CNH_{i,t} + \beta_6 HI_{i,t} + \gamma\ln\left(N_{i,t}\right) \\ &\quad + \delta\ln(G_{i,t}) - \ln\left(\sum_j T_{j,t}\right) \\ &= (1-\rho)E(M_{i,t-1}) + \beta_1 GC_{i,t} + \beta_2 CC_{i,t} + \beta_3 DI_{i,t} + \beta_4 OM_{i,t} \\ &\quad + \beta_5 CNH_{i,t} + \beta_6 HI_{i,t} + \gamma\ln\left(N_{i,t}\right) + \delta\ln(G_{i,t}) + (1 \\ &\quad - \rho)\ln\left(\sum_j T_{j,t-1}\right) - \ln\left(\sum_j T_{j,t}\right). \end{aligned} \tag{F.2}$$

By absorbing the last two terms on the right-hand side of equation (F.2) into the constant term (α) of the regression, and assuming that both terms are time invariant, the following, final specification is obtained as shown in equation (7).

References

BBC Sport. (2014). Olympics and World Cup are the biggest, but what comes next? Retrieved May 02, 2017, from http://www.bbc.com/sport/30326825

Bernard, A., & Busse, M. (2004) Who Wins the Olympic Games: Economic Resources and Medal Totals. The Review of Economics and Statistics, 86:1, 413-417

Boston College. Help bctobit. Retrieved February 20, 2017, from http://fmwww.bc.edu/RePEc/bocode/b/bctobit.html

David Vncent. (2010). "BCTOBIT: Stata module to produce a test of the tobit specification," Statistical Software Components S457163, Boston College Department of Economics.

Engineering Statistics Handbook. Likelihood ratio tests. Retrieved February 20, 2017, from http://www.itl.nist.gov/div898/handbook/apr/section2/apr233.htm

FEI Annual Report 2013. (2016). Retrieved October 06, 2016, from http://inside.fei.org/fei/about-fei/publications/fei-annual-report/2013

Forrest, D., Sanz, I., & J.D. Tena. (2010). Forecasting National Team medal totals at the Summer Olympic Games. International Journal of Forecasting, 26:3, 576-588

Hamilton, J. (1994). Time Series Analysis. Princeton University.

Hoffmann, R., Ging, L., & Ramasamy, B. (2002). Public Policy and Olympic Success. Applied Economics Letters, 9, 545-548.

Honoré, B.E. 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60:3, 533-565

IMF. (2001). Annex to Chapter 6: Classification of the Functions of Government. Retrieved October 10, 2016, from https://www.imf.org/external/pubs/ft/gfs/manual/pdf/ch6ann.pdf

Independent. (2016). Rio 2016: The richest Games in 120 years of Olympic history. Retrieved 02, May, 2017, from http://www.independent.co.uk/news/business/analysis-and-features/rio-2016-olympic-games-richest-ever-usain-bolt-mo-farah-a7171811.html

International Olympic Committee. (2016). Retrieved October 15, 2016, from https://stillmed.olympic.org/Documents/Reference_documents_Factsheets/Women_in_Olympic_Movement.pdf

International Olympic Committee. (2017). Revenue Sources and Distribution. Retrieved 02, May, 2017, from https://www.olympic.org/ioc-financing-revenue-sources-distribution

Investopedia. (2007) Introduction to Stationary and Non-Stationary Processes. Retrieved February 27, 2017, from http://www.investopedia.com/articles/trading/07/stationary.asp

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

Leeds, E., & Leeds, M. (2012). Gold, Silver, and Bronze: Determining National Success in Men's and Women's Summer Olympic Events. Journal of Economics and Statistics, 232, 279-292

Allison, Paul. (2012). When Can You Safely Ignore Multicollinearity? Retrieved February 15, 2017, from http://statisticalhorizons.com/multicollinearity

Söderbom, Mans. (2009). Applied Econometrics Lecture 15: Sample Selection Bias Estimation of Nonlinear Models with Panel Data. University of Gothenburg

Wall Street Journal. (2016). At the Rio Olympics, Women Athletes Bump Against a Gold Ceiling. Retrieved May 02, 2017, from https://www.wsj.com/articles/at-the-rio-olympics-women-athletes-bump-against-a-gold-ceiling-1470425132

Woodridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data (2nd ed.). Massachusetts Institute of Technology

Wooldridge, J. M. (2012). Introductory econometrics: A modern approach (5th ed.). Cengage Learning