Iran. Econ. Rev. Vol. 22, No. 2, 2018. pp. 409-441

Olympic Medals, Economy, Geography and Politics from Sydney to Rio

Martin Grančay*[1], Tomáš Dudáš[2]

Received: September 12, 2017 Accepted: October 2, 2017

Abstract

The paper uses Heckman model to examine the statistical importance of over 140 independent variables on the Olympic performance of all countries participating in the Summer Olympic Games from Sydney 2000 to Rio 2016. We find that countries which export more products have a higher likelihood of winning an Olympic medal than their counterparts exporting fewer products, and explain why this is the case. Statistical importance of gross domestic product per capita, labor force, average temperature and three host effects (previous host, current host, future host) is also confirmed while the role of political variables in Olympic success remains inconclusive.

Keywords: Olympic Games, Medals, Heckman Correction, Economic Variables, Political Variables, Geography.

JEL Classification: L83.

1. Introduction

Not many global events attract as much attention as the Olympic Games. Every four years the best athletes compete in approximately 300 events for glory and medals. They are watched by billions, powered by national pride and the hunger for success. The fastest, strongest and most capable sportsmen become global icons and often serve as role models for youth. As a result, many countries have developed systematic sports policies (Houlihan & Zheng, 2013) to increase or sustain their Olympic performance.

Econometric background of national success at the Olympics has

1. Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia (Corresponding Author: martin.grancay@stuba.sk).

2. Faculty of Economics and Business, Pan-European University, Bratislava, Slovakia (tomas.dudas@paneurouni.com).

been studied at least since Ball's 1972 paper. Factors claimed to have an impact on the number of medals won include economic performance, political situation, geographic location, climate and women empowerment. While there seems to be a broad consensus on the importance of economic and demographic determinants, the story is not so straightforward with the other factors. Researchers found them to be statistically significant or insignificant, depending on the period studied and statistical approach taken. The present article adds to the existing literature by examining over 140 independent variables using data for five Summer Olympic Games from Sydney 2000 to Rio 2016. Unlike some other articles (Forrest et al., 2010) it is not our main aim to develop a model for forecasting; rather we focus on examining the statistical significance of economic, demographic, geographic, climate and political variables and draw conclusions on how countries can achieve better Olympic results. Host-country, ex-host-country and future-host-country effects are included in the analysis as well. To our best knowledge, this is the first research that takes into consideration all of them simultaneously.

The paper is divided into six standard sections. After a brief introduction, a literature review is offered. Section 3 describes sources of data and explains methods used. The results are presented and discussed in section 4, while section 5 offers robustness checks and discussion. Finally, the last section of the paper presents conclusions and ideas for further research.

2. Literature Review

No inquiry into the factors of success at the Olympics would be complete without the 1972 paper of Donald W. Ball. In this paper, Ball offers the first comprehensive analysis of success at the Olympics on a national level using a broad set of geographical, socio-economic and political factors. The result of the analysis show that in order to be successful at the Olympics, nations need to be stable and homogenous in population, literate, modern and "Western", economically prosperous and possess a strong central government. As Ball's paper was published at the height of the cold war, membership in the Communist bloc was also a significant factor of success (as communist countries viewed the Olympics as a possible showcase for propaganda purposes). Grimes et al. (1974) and Levine (1974) published other early studies based on econometric analysis and found that a nation's success at the Olympics is significantly influenced by population size, income per capita, hosting advantage and political system. These studies also demonstrated that communist countries were outliers, as they were winning more medals than their level of economic development and population were likely to predict.

A new wave of studies using more sophisticated statistical methods started to appear after the fall of the Communist bloc in the 1990s. Authors of these studies were using ordered Logit models (Andreff, 2001), an ordered Probit model (Johnson & Ali, 2004) or a Tobit model (Bernard & Busse, 2004). The model created by Bernard and Busse is considered by many to be the best one for estimating the national performance at the Olympics and numerous other studies are based upon it (Jiang & Xu, 2005; Pfau, 2006).

Bernard and Busse demonstrated that the largest part of the Olympic performance on the national level can be predicted by just two variables – population and GDP per capita. Clearly, a larger population means a larger pool of potential athletes to choose from and a high GDP per capita displays the ability to bear the cost to train athletes and send them to the Olympic Games. The results of Bernard and Busse have been successfully confirmed by several subsequent studies (Johnson & Ali, 2004; Andreff, 2013), although studies show that GDP per capita tends to be more important in the case of Winter Games and population in the case of Summer Games (Johnson & Ali, 2004).

However, success at the Olympics is a complex issue and there are also other factors in play. First, the host country advantage has been demonstrated in several studies (Bernard & Busse, 2004) as host countries tend to win more medals than predicted based on their GDP per capita and population. According to Kuper and Sterken (2001) this advantage is visible also in the case of future host countries. For example, Greece doubled its usual amount of medals at the Sydney Olympics (i.e. four years before the Olympics in Athens) and South Korea created a competitive ice hockey team that was able to reach the top division of the Ice Hockey World Championship in 2017 for the first time in the history (just before the Winter Games in Pyeong Chang in 2018). Second, despite the demise of the Communist bloc, models still show the advantages of single party regimes (Johnson & Ali, 2004). Athletes from these countries tend to outperform their counterparts as the training systems of these countries are often harsh. China is probably the best example of these advantages, although North Korea also performs well above expectations at the Summer Games. Finally, other factors that have been suggested to influence Olympic performance include unemployment, ex-host effect (Vagenas & Vlachokyriakou, 2012), gender equality (Berdahl et al., 2015), life expectancy or number of airports (Condon et al., 1999).

3. Data and Methodology

The present paper uses a wide array of variables divided into several groups (Table 1). Dependent variables include total number of medals won at Summer Olympic Games between 2000 and 2016, medal split, medal points counted using the standard approach (3 for gold, 2 for silver, 1 for bronze) and global share of medals and points. We select global share of points as the main dependent variable, and use other indicators to test the model. This is reasonable, because the number of medals awarded changes annually, while the shares remain comparable. The data was taken from the online databank Database Olympics (2011) and updated using BBC (2012, 2016) statistics. Even though we took utmost care while compiling the database, it is possible that it does not reflect all the post-event adjustments in medal counts due to doping scandals. However, as there should not be more than a couple cases which escaped our attention, we do not consider it to be a problem affecting the integrity of results.

Economic variables were taken from UNCTAD's (2017) UNCTADstat Data Center and the World Bank's (2017) World Development Indicators. They include various measures of Gross Domestic Product, export, import, inflation, indicators of foreign direct investment, tariff averages, government consumption etc. Whenever available, variables in both current and constant prices were used. In total 81 economic variables (including same indicators measured differently or logs of indicators) were added into the dataset. Obviously, we never expected to use all of them in the final model, but rather wanted to make sure we have as many variables as possible for correlation analysis which would be the first step of our research.

Demographic variables were provided by UNCTAD (2017) while geographic measures came from the well-known GeoDist database of CEPII (2011), often used by trade economists for gravity modelling. GeoDist provides country-specific geographical variables such as bilateral distances, location-specific dummies (such as landlockedness, continent, official languages etc.) or latitudes and longitudes of capitals and the largest cities.

Several studies have shown that climate affects nations' performance at Olympic Games (Hoffmann et al., 2002; Johnson & Ali, 2004). It is therefore vital to include climate variables such as average annual temperature or average annual precipitation into our analysis. The data was taken from the World Bank (2017). It must be noted that latitude (which we classify as a geographic variable) is also strongly connected to climate.

From among a myriad of potential political variables, we chose to use the Fragile States Index (FSI) and its sub-indices (Fund for Peace, 2017) and Freedom House's (2017) Freedom in the World (FIW) indices. They cover all the countries and territories in the world and are widely accepted. Moreover, the Fragile States Index includes data on 12 sub-indices, which allows us to break political situation in any of the studied countries down into subcategories, such as respect for human rights, demographic pressures, human flight etc. On the negative side, FSI is only available for the last ten years. We also used two dummies to check for the countries' socialist history.

Additional dummy variables were used to take into account the host effect as well as ex-host effect and future-host effect.

After excluding countries or territories which do not have a national Olympic Committee and hence did not participate in any of the studied Olympic Games, we have a set of 200 countries and 5 Olympic Games.[1] The total number of variables at the start of our research reached 7 dependent variables and 144 independent variables. Unfortunately, we do not have data on every single variable

1. The excluded countries/territories are Anguilla, Faeroe Islands, Falkland Islands, French Polynesia, Gibraltar, Greenland, Macau, Montserrat, New Caledonia, Niue, Northern Mariana Islands, Saint Helena, Tokelau, Turks and Caicos, Wallis and Futuna and Western Sahara. Kosovo and independent Olympic athletes were also excluded from the medal counts.

for every single country; hence in some cases our analysis had to be run with a limited dataset. We used a one-year time lag to take into account the fact that Olympic success in a given year is not based on current determinants, but rather is a result of factors originating in previous time periods.[1]

Table 1: Variables and Sources Used

Variable group	Count	Source
Olympic variables (gold, silver, bronze, total, points, shares)	7	Database Olympics (2011). BBC (2012, 2016).
Economic variables (GDP, GDP pc, export, import, trade balance, foreign direct investment flows and stock, trade concentration and diversification, inflation, balance of payments, real effective exchange rate, remittances, tariffs, government expenditure, health expenditure)	81	UNCTAD (2017). World Bank (2017).
Demographic variables (population, density, labor force, urban population, main language)	15	UNCTAD (2017).
Geographic variables (area, internal distance, latitude, longitude, continental dummies)	18	CEPII (2011).
Climate variables (precipitation, temperature)	8	World Bank (2017).
Political variables (fragile states indices, political freedom, post-Soviet dummy, post-Eastern-Bloc dummy)	19	Fund for Peace (2017). Freedom House (2017). Own elaboration.
Additional dummy variables (current, previous and future host countries)	3	Own elaboration.

Source: Own elaboration

As the first step of our research, we ran a simple correlation analysis across the whole dataset to identify mutual relationships between the variables. Conforming to the expectations, many pairs of variables belonging to the same group were found to have high Pearson correlation coefficients, indicating that they should not be made a part of the same regression model. We grouped the variables

1. This effectively means that we used the values of independent variables for 1999, 2003, 2007, 2011 and 2015.

accordingly and prepared a list of variables which could be used in our model without a risk of multicollinearity along with alternative variables where appropriate.[1] Should this step be omitted, we could run into the same problem as Berdahl et al. (2015), who found that gender equality had a positive effect on the number of Olympic medals won, but did not realize their indicator of gender equality correlated with GDP per capita (see Kuppens & Pollet, 2015).

Arguably the most important decision in any research is choosing the right statistical method. Methods used in past studies of determinants of Olympic success vary and include regular OLS (Hoffmann et al., 2002; Johnson & Ali, 2004), Tobit (Bernard & Busse, 2004; Forrest et al., 2010), Heckman correction (partially Johnson & Ali, 2004), Poisson regression (Lui & Suen, 2008; Berdahl et al., 2015), negative binomial regression (Lui & Suen, 2008) or structural equations (Van Tuyckom & Jöreskog, 2012). Each approach has its pros and cons. The main issue to take into account is the "zero medals" problem. From among more than 200 national Olympic committees participating in Rio 2016, less than 50 % won a medal (BBC, 2016). Not only is this share considerably high, but as it is impossible to win a negative number of medals, zero is the lower limit for the variable, which makes OLS unusable. One of the approaches that is sometimes taken is omitting unsuccessful countries from the analysis; however, this would quite obviously lead to a bias in results, because the zeroes are not randomly distributed. This is a similar problem which is present in gravity models of trade (see for example Head & Mayer, 2014).

A simple solution which immediately comes to mind is using Poisson regression. Given the method's suitability for count data, this could seem appropriate. However, using Poisson regression for datasets with a large share of zeroes is not recommended (Burger et al., 2009). Moreover, we prefer the global share of medal points as a dependent variable, instead of simple count data with medals or points. This is similar to what has been done by Bernard and Busse (2004). As a result, Tobit and Heckman correction become the most viable alternatives. If sufficient care is taken and one corrects for heteroscedasticity and

1. For example, population, labor force and their logs are highly correlated, hence they are alternatives to each other.

autocorrelation in the standard errors, Tobit is a reasonable choice (Bernard & Busse, 2004). Nevertheless, in the present paper we choose Heckman correction and keep Tobit as one of our robustness checks.

The use of Heckit, as the method is often called, brings about two main advantages. First, it minimizes the problem of heteroscedasticity. Second, it allows us to separately estimate factors that decide whether a country wins any Olympic medal and factors that determine the number of medals. On the other hand, Heckman correction also has several problems. These include the difficulty of identifying correct selection regressors and possible issues with non-positive-definite covariance matrices. As we shall see in the next sections, this happens only rarely in our case. If maximum likelihood estimation is not possible, we use the 2-step model.

4. Results and Interpretation

Before proceeding to the regression analysis itself, it is useful to prepare a correlation matrix of all the variables included in the dataset. It allows us to identify mutual relationships between the variables and offers a simple preliminary look of how strong the link between the dependent and independent variables is. The highest correlation coefficients can be witnessed with economic variables, in multiple cases exceeding 0.7 (Table 2). Government expenditures, GDP, trade flows and foreign direct investment stock/flows appear to be the most important of them.[1] From among other variables, area, internal distance (average distance between producers and consumers in a country), labor force and population have the highest correlation coefficients. Climate and political variables show a weak negative correlation with Olympic variables. This is in line with literature and indicates that less democracy and colder climate might bring additional Olympic medals (Hoffmann et al., 2002; Andreff, 2013). We will take all these facts into account later when designing the regression model. For now, the main purpose of this exercise was to find mutual relationships between the dependent variables to avoid

---

1. We have to stress that these are the results of a simple correlation without controlling for other factors, so they cannot lead to any serious conclusions. They might only show a certain pattern.

problems with multicollinearity in further steps of our analysis.[1] We clustered variables into several groups according to their level of correlation (Table 3).[2] Variables in the same group have mutual correlation higher than 0.60 and should not be used in the same regression analysis. Variables not reported in the table do not exhibit high correlation with any other dependent variables.

Table 2: Highest Correlation Coefficients of Independent Variables with Olympic Variables

	Gold	Silver	Bronze	Total	Points	Medals´ share	Points´ share
			Economic variables				
Gov. expenditures (constant 2005)	0.781	0.786	0.773	0.805	0.804	0.802	0.802
Gov. exp. (current)	0.765	0.767	0.749	0.785	0.786	0.780	0.782
GDP (constant 2005)	0.768	0.758	0.731	0.778	0.781	0.775	0.779
Import	0.768	0.755	0.735	0.778	0.781	0.772	0.777
GDP (current)	0.765	0.751	0.728	0.773	0.777	0.768	0.773
Remittances (paid)	0.737	0.733	0.698	0.743	0.748	0.738	0.744
Export	0.726	0.701	0.710	0.736	0.737	0.730	0.732
FDI (inward stock)	0.677	0.686	0.644	0.691	0.694	0.685	0.690
FDI (outward stock)	0.664	0.681	0.649	0.686	0.686	0.680	0.682
FDI (inward flows)	0.643	0.621	0.603	0.643	0.648	0.639	0.645
FDI (outward flows)	0.612	0.612	0.603	0.629	0.628	0.625	0.626
ln Gov. expenditures (constant 2005)	0.547	0.605	0.610	0.604	0.591	0.604	0.591
ln Remittances (paid)	0.531	0.576	0.575	0.576	0.566	0.574	0.565
ln Gov. exp. (current)	0.512	0.568	0.572	0.567	0.554	0.565	0.553
ln GDP (constant 2005)	0.489	0.542	0.548	0.542	0.530	0.542	0.530
ln GDP (current)	0.479	0.531	0.539	0.532	0.519	0.530	0.519

1. Even though Heckit performs better with multicollinearity than OLS and many other models, this still does not mean that care should not be taken to limit its presence.
2. Due to its size, we do not report the whole correlation matrix here.

Table 2: Highest Correlation Coefficients of Independent Variables with Olympic Variables

	Gold	Silver	Bronze	Total	Points	Medals´ share	Points´ share
Demographic variables							
Labor force	0.540	0.410	0.416	0.474	0.495	0.473	0.494
Population	0.482	0.374	0.382	0.429	0.446	0.428	0.445
World population share	0.482	0.373	0.378	0.428	0.445	0.428	0.445
Geographic variables							
Area	0.629	0.644	0.669	0.668	0.659	0.668	0.659
Internal distance	0.536	0.558	0.574	0.574	0.565	0.574	0.565
Latitude (absolute)	0.313	0.360	0.384	0.362	0.349	0.363	0.349
Climate variables							
Avg. min. temperature	-0.345	-0.378	-0.414	-0.391	-0.377	-0.391	-0.377
Average temperature	-0.344	-0.378	-0.416	-0.391	-0.377	-0.391	-0.377
Avg. max. temperature	-0.332	-0.367	-0.405	-0.379	-0.365	-0.379	-0.365
Political variables							
FSI external intervention	-0.344	-0.407	-0.394	-0.392	-0.382	-0.392	-0.382
FSI human flight	-0.290	-0.375	-0.360	-0.350	-0.336	-0.349	-0.336
FSI poverty	-0.297	-0.358	-0.359	-0.347	-0.335	-0.347	-0.335

Notes: Sorted by absolute value of correlation with the "Points' share" variable. All variables with correlation coefficient higher than 0.60 or lower than -0.60, and/or the leading three performers of each group of variables.
FSI: Failed / Fragile States Index.

Not surprisingly, many economic indicators – such as GDP, trade flows, FDI, government expenses etc. – correlate with each other. This is in line with general economic knowledge. The same can be said about alternative indicators measuring similar categories, such as various political indicators, various temperature variables or various tariff variables. A more interesting discovery is the high correlation between GDP per capita and the Fragile States Index, or between tariff variables and average temperature. Both of them can be readily explained, but were probably not expected.

It goes without saying that being in different groups is not a sufficient condition for variables to be used in the same regression model.

Table 3: Groups of Variables with Mutual Correlation Higher Than 0.60 or Lower than -0.60

Group	Variables
1	GDP, trade flows, indicators of FDI, remittances paid, government expenses, health expenses
2	ln GDP, ln labor force, ln population, ln trade flows, exported products, imported products, import diversity, ln indicators of FDI, ln remittances received, ln remittances paid, ln government expenses, ln health expenses, ln area, ln internal distance, some FSI indicators
3	GDP pc, ln indicators of FDI, FSI indicators
4	ln GDP pc, urban population share, ln indicators of FDI, latitude, ln remittances paid, FSI indicators
5	labor force, population, global population share, remittances received
6	trade balance, balance of payments, FDI stock
7	all tariff variables, average temperature, some FSI indicators
8	all GDP growth variables
9	latitude of the capital city, latitude of the largest city, their absolute values (=distance from the equator)
10	population density, labor force density
11	area, ln area
12	all temperature variables
13	all political indicators

A crucial task for performing Heckit correctly is identifying suitable selection regressor(s). Selection regressors explain why sometimes the dependent variable has a non-zero value, and sometimes it is zero or unobserved. In our case selection regressor(s) should explain why some countries win Olympic medals whereas others do not. The main problem consists in the requirement that selection regressor should be a variable which is highly correlated with the selection dummy, but does not directly influence the dependent variable. This means we are searching for a variable connected to whether country wins a medal, but not influencing the number of medals. A simple correlation analysis detects two possibilities: (1) the number of products a country exports and (2) absolute value of the largest city's latitude. The correlation coefficients reach 0.621 with the selection variable (medal dummy) and 0.364 with the dependent variable for the former and 0.513 and 0.345 for the latter.

The importance of latitude is not unexpected. Latitude is one of the possible proxies for climate and several papers have shown that climate is a statistically significant factor in determining the number of Olympic medals (Hoffmann et al., 2002; Johnson & Ali, 2004 etc.). It is plausible that its role is even higher when it comes to our selection variable.

Much more surprising is the possible role of the number of products exported.[1] At first thought the indicator has no link to Olympic medals. However, one has to consider that countries which export the most products tend to have large economies, while countries with very concentrated exports usually have small economies.[2] Number of products exported therefore seems to be a pretty good proxy of GDP. In addition, consider that almost everyone who made research on determinants of Olympic medals has come to the conclusion that population and GDP per capita are statistically significant factors. Knowing that simple multiplication of the two gives GDP, it becomes obvious that number of products exported – as a proxy for GDP – indeed has a connection with Olympic medals. Most importantly for us, the correlation with the number of medals is very weak, while with r=0.621 the variable is the best predictor of medal dummy in our dataset. Figure 1 uses data from Rio 2016 to illustrate how the number of countries winning at least a medal increases with the number of products exported.

1. Data is taken from UNCTAD (2017) and is at the three-digit SITC, Rev. 3 level. Maximum value is 261.
2. Correlation between ln GDP and the number of products exported, r=0.865.

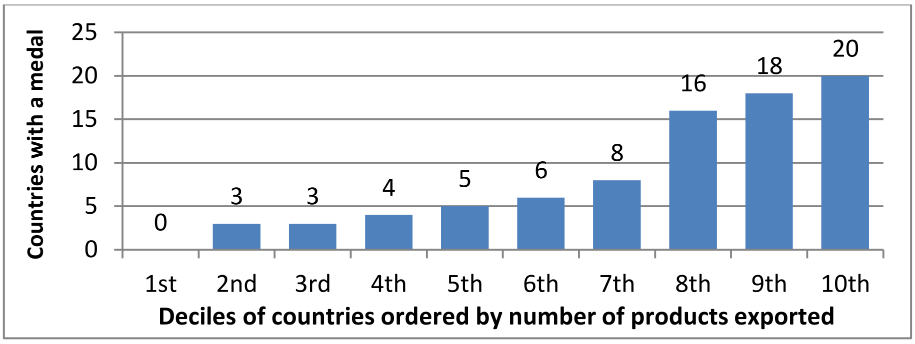

Figure 1: Countries with at Least one Medal from Rio 2016 By Deciles Ordered by Number of Products Exported
Source: Own Elaboration Based on UNCTAD (2017) and BBC (2016)

Both above-mentioned selection variables will be used in the first phase of our analysis. Yet designing a suitable selection equation is only a part of the story. An equally important task is estimating the main regression equation. We will do this by combining results from Table 2 with approaches of previous researches of this topic. The vast majority of papers published to date includes population and GDP per capita among dependent variables and find they are both statistically significant. This has logic as the first is a good proxy for the size of the pool of possible sportsmen in a country and the second is a direct measure of welfare and an indirect measure of financial resources available to sportsmen. We will use GDP per capita in constant 2005 US dollars throughout the paper. However, instead of using population, we opt to use labor force in thousands. This is defined as "people aged 15 and older" (UNCTAD, 2017) and according to our opinion is a better proxy for the size of the pool of possible sportsmen than population which also includes babies and children. Other researchers have mostly used it in percentage terms (Vagenas & Vlachokyriakou, 2012) or partially as a share of women employees within the Gender Inequality Index (Lowen et al., 2016). Both labor force and GDP per capita was log transformed.

Other variables that cannot be omitted from the model relate to the host effect. A country hosting Olympic Games has a number of advantages which it can transform into additional medals. These include lower costs of participation for domestic athletes, higher motivation in front of the home crowd or the possibility of host countries to tailor facilities to meet the needs of their sportsmen (Bernard & Busse, 2004).

Researchers have also confirmed the importance of ex-host effect (Hoffmann et al., 2002; Vagenas & Vlachokyriakou, 2012) and future-host effect (Forrest et al., 2010). All of these variables have a binary form. The host variable takes the value 1 only for the country hosting the Olympic Games in question; the ex-host variable takes the value 1 for the country hosting the previous Olympic Games and the future-host variable is 1 for the country due to organize the next Summer Olympics.[1] The ex-host effect can be reasoned by the existence and quality of sporting and sport-related infrastructures in former host countries as well as by supposed public and governmental support for sports which is clearly necessary for organizing a successful event (Hoffmann et al., 2002). Even if the support comes to a sudden stop after the Games are over, a certain level of inertia can be expected. Forrest et al. (2010) explain the future-host effect by assuming that "a country will wish to do particularly well in terms of medals when it stages the Games", and so the government will begin preparing as soon as possible by investing more financial resources into elite athletes. Considering that the Games are usually awarded seven years in advance, the effect of these preparations can already be visible four years before the event.

So far the selection of variables has been pretty straightforward. However, political, climate, geographic and additional economic factors that might influence the number of Olympic medals is much more difficult to identify. We have therefore run dozens of regression analyses to find the best options. We have concluded that Freedom House's (2017) Freedom in the World (FIW) – Civil Liberties index appears to be the best option for a political variable, while latitude or average temperature have the highest statistical significance from among the geographic and climate variables. No other economic variables that had not been excluded by correlation analysis presented in Table 3 (such as inflation, real effective exchange rate or population density) were found to have a significant effect on the number of medals.

---

1. In other researchers' works the ex-host variable often takes into account not only the previous Olympic Games, but the whole history of Olympic hosting (see for example Hoffmann et al., 2002). While this certainly is a valid possibility, we choose a different path.

Table 4: Basic Heckman Models for Summer Olympic Games 2000-2016

	Model 1 (Pts. share)	Model 2 (Pts. share)	Model 3 (Pts. share)	Model 4 (Pts. share)
Main regression. Dependent variable: Global share of medal points in %				
Intercept	-12.927*** (0.806)	-12.034*** (0.829)	-11.849*** (0.759)	-10.356*** (0.739)
ln LabFor	0.740*** (0.044)	0.605*** (0.048)	0.732*** (0.042)	0.618*** (0.042)
ln GDPpc	0.596*** (0.054)	0.528*** (0.052)	0.656*** (0.051)	0.659*** (0.048)
D_Host	2.204*** (0.470)	2.004*** (0.388)	2.235*** (0.420)	2.247*** (0.319)
D_PrevHost	2.156*** (0.571)	1.622*** (0.504)	1.874*** (0.526)	1.410*** (0.438)
D_NextHost	1.095** (0.515)	1.090** (0.464)	1.190** (0.475)	1.382*** (0.419)
Civil liberties	0.066 (0.041)	0.053 (0.039)	0.093** (0.041)	0.082** (0.41)
Latitude (abs. value)	0.026*** (0.004)	0.055*** (0.005)	-	-
Avg. temperature	-	-	-0.051*** (0.006)	-0.082*** (0.007)
Lambda	1.634*** (0.078)	1.713*** (0.077)	1.659*** (0.076)	1.700*** (0.073)
Selection equation. Dependent variable: Dummy for medals				
Intercept	-2.093*** (0.131)	-2.422*** (0.141)	-2.034*** (0.130)	-2.320*** (0.146)
Exported products	0.011*** (0.001)	0.008*** (0.001)	0.011*** (0.001)	0.009*** (0.001)
Latitude (abs. value)	-	0.029*** (0.003)	-	0.022*** (0.002)
Model characteristics				
Total observations	971	971	963	963
Censored observations	572	572	572	572
Year dummies	Yes	Yes	Yes	Yes

Notes: Own calculations. Standard errors in parentheses. Year dummies not reported.
* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %.

The results of the four most promising basic Heckman models for

Summer Olympic Games 2000-2016 are reported in Table 4. We use exported products or exported products and latitude as selection regressors, and alternate latitude and average temperature as climate variables in the main regression. All the models include year dummies.

Coefficients in all four selection equations have the expected signs and are highly statistically significant. Countries which export more products have a higher likelihood of winning an Olympic medal than their counterparts exporting fewer products. In other words, the more diversified export structure the higher the chance to win a medal. Obviously, this does not mean that ministries of sports should now try to influence the national export structure in hopes of the Olympic success. As we explained before, the number of products exported can be considered a proxy for the country's wealth, for GDP. As richer countries tend to invest more in sports (Andreff, 2001), higher probability of winning at least one Olympic medal is just a logical consequence. The positive coefficient of latitude (expressed as an absolute value) is in line with literature and indicates that countries in milder and colder climates are more likely to score a medal than countries close to the equator.

Most importantly, coefficients in the main regression equations also have the anticipated signs and with a minor exception are all statistically significant. Labor force and GDP per capita have a positive effect on the number of medals, confirming that countries with a larger population in productive age have a better chance of generating multiple top-class sportsmen, and athletes from countries with a higher standard of living achieve better Olympic results.

An interesting comparison can be drawn among the three host effects. All of them are highly statistically significant, and the magnitude of the influence they have on the number of medals won follows the order current host – former host – future host. A country which hosts the Olympic Games earns more extra medals than the ex-host country as well as more extra medals than the country preparing to host the next Summer Olympics. Moreover, former host countries win more extra medals than future host countries. These findings indicate that inertia from substantial investment into sports required to organize Olympic Games lasts longer than one Olympic cycle and is stronger than the effect of preparation for the next games.

Both average temperature and latitude as our interchangeably used climate variables appear to fit well with our expectations. Similar to the latitude's role in the selection equation, milder and colder countries win more medals than their warmer counterparts. It might be argued however that these variables should not have a linear effect on the Olympic success, but rather a quadratic one as demonstrated by Hoffmann et al. (2002) – countries which are either extremely hot or extremely cold win fewer medals. Despite using several variables to test this in separate regressions (including temperature difference from world average and quadratic versions of both latitude and temperature), we were not able to verify this claim. We believe the linearity of the relationship is statistically substantiated by a large number of countries in warm climates which win relatively few medals and a large number of countries in mild climates which win relatively many medals. The number of countries located in cold climates is low and their Olympic performance does not negatively affect statistical significance of this relationship, especially considering they include Russia and Canada, both being very successful Olympic countries (Figure 2).

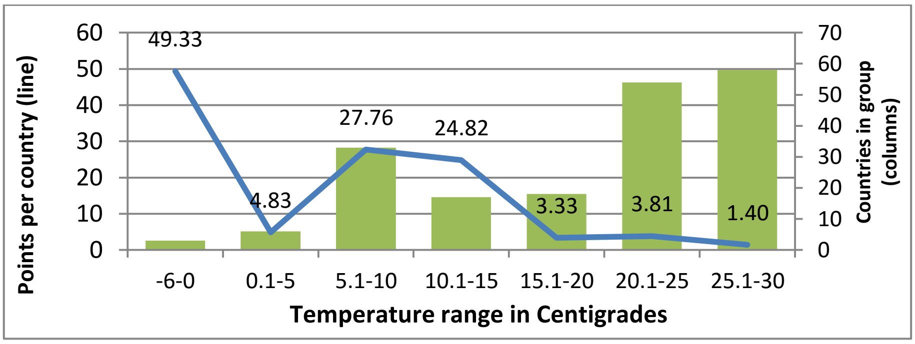

Figure 2: Medal Points Won in Rio 2016 per Country by Temperature Range
Source: Own elaboration based on UNCTAD (2017) and BBC (2016)

The political variable is positive in all four models, but statistically significant only in those where the average temperature is used as a measure of climate. Countries with lower respect for civil liberties (which translates into a higher score of the index) tend to win more Olympic medals than democratic countries respecting personal freedoms. In the past this used to be explained by the "Soviet effect"

(Bernard & Busse, 2004), arguing that central planning of the countries in the Soviet sphere of influence allowed for more specialization and “more national resources were dedicated to training and supporting athletes […] than in market-based economies” (Van Tuyckom & Jöreskog, 2012). The latter argument can probably still be used for many non-democratic countries with less respect for civil liberties, who might, for example, single out future athletes at an early age and subject them to rigorous specialist training (Forrest et al., 2010).

All the models have statistically significant lambda coefficients indicating there is a sample selection bias and Heckit should be preferred over OLS. Log-likelihoods are in the range of -1022 to -1090. While the models are similar to each other, we will use model 3 as the basis for further research due to the fact that it appears to be the most robust one based on the results of robustness tests presented in section 5. This model works well also with other dependent variables, such as total number of medals, total medal points, number of gold medals, or share of medals (Table 5). All the dependent variables remain statistically significant, have the same sign and can be interpreted in the same way as in the basic model.

Table 5: Heckman Models with Alternative Dependent Variables, Summer OG 2000-2016

	Model 5 (Total points)	**Model 6 (Total medals)**	**Model 7 (Gold)**	**Model 8 (Medals’ share)**
Main regression based on model 3. Dependent variable: Various				
Intercept	-219.422*** (14.093)	-107.375*** (6.828)	-39.949*** (2.728)	-11.347*** (0.719)
ln LabFor	13.587*** (0.779)	6.729*** (0.378)	2.420*** (0.152)	0.709*** (0.040)
ln GDPpc	12.155*** (0.949)	5.979*** (0.461)	2.171*** (0.186)	0.631*** (0.049)
D_Host	41.577*** (7.801)	18.935*** (3.814)	8.866*** (1.570)	1.988*** (0.401)
D_PrevHost	34.882*** (9.775)	16.636*** (4.679)	7.227*** (2.002)	1.746*** (0.492)
D_NextHost	22.139** (8.872)	11.367** (4.484)	4.067** (1.685)	1.199** (0.470)

Table 5: Heckman Models with Alternative Dependent Variables, Summer OG 2000-2016

	Model 5 (Total points)	Model 6 (Total medals)	Model 7 (Gold)	Model 8 (Medals' share)
Civil liberties	1.716** (0.757)	0.867** (0.368)	0.312** (0.147)	0.092** (0.039)
Avg. temperature	-0.944*** (0.118)	-0.485*** (0.058)	-0.160*** (0.023)	-0.051*** (0.006)
Lambda	30.774*** (1.406)	14.888*** (0.678)	5.881*** (0.282)	1.571*** (0.071)
Selection equation. Dependent variable: Dummy for medals				
Intercept	-2.036*** (0.130)	-2.040*** (0.130)	-2.048*** (0.131)	-2.036*** (0.130)
Exported products	0.011*** (0.001)	0.011*** (0.001)	0.011*** (0.001)	0.011*** (0.001)
Model characteristics				
Total observations	963	963	963	963
Censored observations	572	572	572	572
Year dummies	Yes	Yes	Yes	Yes

Notes: Own calculations. Standard errors in parentheses. Year dummies not reported.
* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %.

Assessing the role of political factors in Olympic success is tricky. Numerous studies have confirmed the significance of the "Soviet effect" in postwar Olympic Games. However, the fall of the iron curtain in 1989 brought about new socio-economic and political reality, and the number of medals won by post-Soviet countries decreased immediately in Barcelona 1992 (Shughart & Tollison, 1993). Since then the results have been inconclusive: some researchers find the political variable significant (Bredtmann et al., 2016), while others do not (Lowen et al., 2016) and some argue the change is underway (Forrest et al., 2010). The reason might be different time periods taken into account, different methods or different political variables used (planned economy dummy, Polity indicator, post-Soviet dummy etc.). To see how the latter affects our results, we have run basic model regressions with all the other political variables included in our dataset (Table 6).

When Freedom in the World – Political Rights Index is used instead of the Civil Liberties Index (model 9), the results remain virtually unchanged. The same is true for the dummy for "free countries" as defined by FIW (model 10), even though here the coefficient is negative due to the reverse construction of the dummy. Two additional political variables – a dummy for post-Eastern-Bloc countries and post-socialist countries (models 11-12)[1] – also lead to similar results. Finally, we test the significance of the Fragile States Index (model 13), and despite having to limit the dataset to the last three Summer Olympic Games, all the conclusions remain valid.[2] It therefore appears that the model is robust to different specifications of the political variable and that political factors still have an important effect on Olympic success.

Table 6: Heckman Models with Alternative Political Variables, Summer OG 2000-2016

	Model 9 (FIW-PR)	**Model 10 (FIW-Free)**	**Model 11 (D_PostEB)**	**Model 12 (D_PostSoc)**	**Model 13 (FSI)**
Main regression based on model 3. Dependent variable: Global share of medal points in %					
Intercept	-11.791*** (0.751)	-11.307*** (0.672)	-12.951*** (0.750)	-14.505*** (0.822)	-13.053*** (1.423)
ln LabFor	0.738*** (0.042)	0.736*** (0.042)	0.772*** (0.042)	0.777*** (0.040)	0.716*** (0.056)
ln GDPpc	0.648*** (0.050)	0.640*** (0.048)	0.708*** (0.046)	0.837*** (0.053)	0.756*** (0.105)
D_Host	2.240*** (0.423)	2.264*** (0.417)	1.996*** (0.353)	1.893*** (0.325)	2.540*** (0.604)

1. The first dummy includes only countries that formed a part of the Soviet Union or of the Eastern Bloc in Europe, whereas the second one includes socialist countries from around the world.
2. Additionally, we have run separate regression analyses for every single sub-index of the FSI. The following indices were found to be statistically significant: FSI-Legitimacy of the State, FSI-Human Rights, FSI-Uneven Development, FSI-Poverty and Economic Decline and FSI-Factionalized Elites. The following indices were found not to be statistically significant: FSI-Demographic Pressures, FSI-Refugees and IDPs, FSI-Group Grievance, FSI-Human Flight, FSI-Public Services, FSI-Security Apparatus and FSI-External Intervention.
It needs to be noted that FSI indicators exhibit high levels of correlation with ln GDP per capita (correlation coefficients are between 0.55 and 0.80) and are therefore not ideal choices for the model.

Table 6: Heckman Models with Alternative Political Variables, Summer OG 2000-2016

	Model 9 (FIW-PR)	**Model 10 (FIW-Free)**	**Model 11 (D_PostEB)**	**Model 12 (D_PostSoc)**	**Model 13 (FSI)**
D_PrevHost	1.938*** (0.528)	1.894*** (0.521)	1.505*** (0.451)	1.222*** (0.439)	1.324** (0.599)
D_NextHost	1.187** (0.482)	1.236*** (0.477)	1.100** (0.435)	1.037** (0.416)	0.850 (0.677)
Political variable	0.075** (0.034)	-0.311** (0.145)	0.919*** (0.143)	1.186*** (0.152)	0.014** (0.007)
Avg. temperature	-0.050*** (0.006)	-0.050*** (0.006)	-0.029*** (0.007)	-0.016** (0.007)	-0.060*** (0.010)
Lambda	1.652*** (0.076)	1.665*** (0.075)	1.699*** (0.071)	1.677*** (0.070)	1.657*** (0.101)
Selection equation. Dependent variable: Dummy for medals					
Intercept	-2.030*** (0.130)	-2.023*** (0.130)	-2.036*** (0.127)	-1.932*** (0.133)	-2.066*** (0.183)
Exported products	0.011*** (0.001)	0.011*** (0.001)	0.011*** (0.011)	0.010*** (0.001)	0.011*** (0.001)
Model characteristics					
Total observations	963	963	965	965	579
Censored observations	572	572	572	572	336
Year dummies	Yes	Yes	Yes	Yes	Yes

Notes: Own calculations. Standard errors in parentheses. Year dummies not reported. Political variables used: FIW-PR – Freedom in the World Political Rights (best 1-worst 7), FIW-Free – Freedom in the World free countries dummy (not free 0-free 1), D_PostEB – dummy for post-Eastern-Bloc countries (no 0-yes 1), D_PostSoc – dummy for post-socialist countries (no 0-yes 1), FSI – Failed/Fragile States Index (best 0-worst 120). FSI available only for the last three Olympic Games.
* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %.

Similar results are achieved in all basic and alternative models when share on global population is used instead of ln LabFor.

5. Robustness Checks and Discussion

To test the robustness of our model, we have followed the usual practice and have run multiple regressions on subsets of the main dataset. When 100 richest countries by GDP per capita are considered (model 14), the dummy for future host loses statistical significance (p-value = 0.20), otherwise there are no big differences from the basic model. Models limited to the countries from the Northern hemisphere (15) or those with direct access to the sea (16) are similar to the basic model as well, with all variables being significant, having the same sign and order of coefficients.[1]

An important test consists in excluding outliers, defined as all values of dependent variable higher/lower than the mean ± 2 standard deviations, from the dataset (model 17). It is well known that the presence of outliers can affect the performance of regression models. In our case, 39 observations were dropped. These included mostly USA, Russia, China, Japan, France, Germany and Australia. As there is a lower limit on the number of medals a country can win (zero), there were no lower-bound outliers. Due to the fact that 50 % of medal points were dropped this way, it is not surprising that the values of all coefficients decreased. Civil liberties and more importantly even ex-host effect became statistically insignificant. We believe the differences from the basic model are mainly caused by the fact that ten of the dropped observations were host, ex-host or future host countries. Importantly, after excluding outliers only a single observation has a non-zero value of the ex-host dummy (Greece). A very poor performance of Greece in Beijing 2008 (4 medals compared to 16 in Athens 2004) explains statistical insignificance of the D_PrevHost variable. Once host country observations are reintroduced into the analysis, all the independent variables are significant again.

Finally, we run a Tobit regression (model 18) using variables from the main equation. Coefficients remain significant; however, the ex-host effect now appears to be more important than the host effect. This has no logical explanation and is an indication that Heckman regression might be more appropriate. OLS regression with heteroscedasticity-robust standard errors generates similar results with insignificant civil liberties variable.

1. We did not run separate regressions for countries from the Southern hemisphere or landlocked countries as these constitute only approx. 20 % of the total.

Table 7: Robustness Checks (Heckman), Summer OG 2000-2016

	Model 14 (100 richest)	Model 15 (North)	Model 16 (sea)	Model 17 (w/o outliers)	Model 18 (Tobit)
Main regression based on model 3. Dependent variable: Global share of medal points in %					
Intercept	-10.984*** (1.310)	-11.813*** (0.820)	-11.923*** (1.032)	-4.187*** (0.359)	-11.553*** (1.207)
ln LabFor	0.902*** (0.057)	0.744*** (0.045)	0.782*** (0.050)	0.281*** (0.021)	0.827*** (0.084)
ln GDPpc	0.437*** (0.102)	0.647*** (0.055)	0.647*** (0.071)	0.248*** (0.024)	0.632*** (0.070)
D_Host	1.798*** (0.398)	2.585*** (0.547)	2.485*** (0.529)	1.265*** (0.395)	3.657*** (1.150)
D_PrevHost	1.347*** (0.440)	1.639*** (0.533)	2.387*** (0.634)	-0.233 (0.557)	3.832*** (1.083)
D_NextHost	0.621 (0.487)	1.584*** (0.522)	1.146** (0.567)	0.856** (0.395)	1.271** (0.619)
Civil liberties	0.089 (0.062)	0.107** (0.044)	0.099* (0.052)	0.008 (0.019)	0.091* (0.047)
Avg. temperature	-0.054*** (0.008)	-0.055*** (0.007)	-0.066*** (0.009)	-0.021*** (0.003)	-0.087*** (0.011)
Lambda	1.834*** (0.088)	1.701*** (0.082)	1.669*** (0.095)	0.711*** (0.037)	-
Selection equation. Dependent variable: Dummy for medals					
Intercept	-2.488*** (0.230)	-1.845*** (0.140)	-2.430*** (0.178)	-1.888*** (0.130)	-
Exported products	0.014*** (0.001)	0.010*** (0.001)	0.013*** (0.001)	0.010*** (0.001)	-
Model characteristics					
Total observations	476	757	783	924	869
Censored observations	200	413	460	572	478
Year dummies	Yes	Yes	Yes	Yes	Yes

Notes: Own calculations. Standard errors in parentheses. 100 richest countries based on constant GDP per capita in 2016. North is based on latitudinal position of the countries' largest cities. Model 16 includes all non-landlocked countries. Outliers were defined as all values of dependent variable higher/lower than mean ± 2 standard deviations. Tobit – QML robust standard errors.

* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %

Table 8: Heckman Models by Olympic Games, Summer OG 2000-2016

	Model 19 (OG 2000)	Model 20 (OG 2004)	Model 21 (OG 2008)	Model 22 (OG 2012)	Model 23 (OG 2016)
Main regression based on model 3. Dependent variable: Global share of medal points in %					
Intercept	-9.386*** (1.261)	-16.713*** (2.919)	-7.913*** (1.797)	-9.833*** (1.655)	-10.692*** (1.639)
ln LabFor	0.619*** (0.073)	1.034*** (0.145)	0.571*** (0.095)	0.630*** (0.096)	0.663*** (0.095)
ln GDPpc	0.521*** (0.084)	0.907*** (0.193)	0.419*** (0.115)	0.549*** (0.113)	0.591*** (0.114)
D_Host	5.522*** (1.465)	0.691 (1.406)	9.718*** (1.740)	5.165*** (1.299)	0.812 (1.308)
D_PrevHost	7.230*** (1.504)	3.994*** (1.529)	-0.126 (1.603)	7.267*** (1.368)	5.286*** (1.362)
D_NextHost	0.952 (1.302)	2.541 (1.617)	3.243** (1.637)	0.088 (1.230)	1.122 (1.217)
Civil liberties	0.102 (0.067)	0.242** (0.122)	-0.081 (0.104)	-0.021 (0.088)	0.058 (0.088)
Avg. temperature	-0.047*** (0.011)	-0.079*** (0.020)	-0.038*** (0.010)	-0.038*** (0.014)	-0.047*** (0.014)
Lambda	1.434*** (0.129)	1.821*** (0.490)	1.585*** (0.144)	1.523*** (0.160)	1.605*** (0.162)
Selection equation. Dependent variable: Dummy for medals					
Intercept	-1.739*** (0.243)	-2.473*** (0.306)	-1.733*** (0.293)	-1.994*** (0.338)	-2.054*** (0.317)
Exported products	0.010*** (0.001)	0.013*** (0.002)	0.009*** (0.002)	0.010*** (0.002)	0.010*** (0.002)
Model characteristics					
Total observations	191	193	193	192	194
Censored observations	114	122	109	113	114

Notes: Own calculations. Standard errors in parentheses. For OG 2004 MLE model not possible (matrix not positive definite), two-step Heckit used instead.
* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %.

All in all, the tests appear to show that the basic model is robust and correctly specified.[1] The main problem might be the political

1. We ran the robustness checks for all basic models (models 1-4) in the earlier stages of our research and partially based on their results 3 was selected as the best model.

variable, which is not significant in some cases. This becomes even more apparent, when separate analyses are run for each Olympics in the studied period (Table 8).

Considering the role of luck, one-off successes, one-off failures and a smaller number of observations, it is not surprising that models for each Olympics are different; several things that all of them have high significance of the demographic and economic variable as well as average temperature. Conversely, dummies for the three host effects show mixed results. For example, the host effect is not significant in 2004 and 2016, while the ex-host effect is not significant in 2008. Taking into account that these relate to the Olympics that took place in Greece and Brazil, a possible explanation is that small countries (either by the size of the population or by GDP per capita) are not able to take full advantage of the host effect. However, testing this possibility would require considering a much longer time period than in the current paper. What appears clear though is that the future-host effect is not significant on the individual-Games level, although a pattern of lower p-value for bigger countries can be detected.

With a single exception of 2004, the political variable is not significant. This is contrary to what the main model and its alternatives with other political variables show. A possible reason is the relatively small number of observations, with a maximum of 194 for each Olympics. On the other hand, it cannot be ruled out that political factors really ceased to be important after the fall of the Iron Curtain, as hinted by Lowen et al. (2016), but our main model still finds them to be significant due to some statistical issues. For example, there is a certain level of correlation between civil liberties and ln GDP per capita (r=-0.56), which might be influencing the results, even though neither variance inflation factors nor Belsley-Kuh-Welsch condition numbers indicate problems with collinearity. The main model is based on a much larger sample; hence possible collinearity should have relatively little effect (Mason & Perreault Jr., 1991), arguably smaller than in individual models. Even though this leads us to incline to the main model interpretation, we should probably take a conservative approach and state that the role of the political variable is far from certain.

Models based on continental data are also comparatively divergent

(Table 9). Because of statistical limitations of Heckit, only the models for Europe, Asia and Americas can be presented. For Americas, the only significant regressors are labor force and GDP per capita. An obvious reason here is the very high share of US medals (55 %) and medal points (58 %), which outweigh all other observations and disallow other variables to gain statistical significance. In the Asian model, lambda is not significant, meaning that ordinary OLS should yield better results; yet all the coefficients are in line with expectations. An interesting feature of the European model is the negative sign of the civil liberties coefficient, indicating that on the old continent democratic countries win more medals. This can be an accidental result caused by high correlation between the GDP per capita and FIW-Civil Liberties index in Europe (r=-0.75). However, it can also be a sign that once countries reach a certain level of democracy, any further increases in the level of freedom will foster competition and lead to better Olympic performance. If this is the case quadratic rather than linear function might be a better choice for the variable.

Table 9: Heckman Models by Region, Summer OG 2000-2016

	Model 24 (Europe)	**Model 25 (Asia)**	**Model 26 (Americas)**
Main regression based on model 1. Dependent variable: Points			
Intercept	-6.824*** (0.803)	-11.141*** (1.548)	-36.001*** (7.602)
ln LabFor	0.893*** (0.064)	0.716*** (0.071)	1.569*** (0.359)
ln GDPpc	0.091* (0.054)	0.820*** (0.085)	2.112*** (0.578)
D_Host	1.959*** (0.300)	6.420*** (1.038)	-1.481 (3.281)
D_PrevHost	0.788*** (0.303)	4.456*** (1.035)	2.843 (3.360)
D_NextHost	1.782** (0.703)	1.306* (0.732)	-2.461 (3.263)
Civil liberties	-0.199*** (0.062)	0.284*** (0.064)	-0.015 (0.275)
Avg. temperature	-0.048** (0.023)	-0.158*** (0.011)	0.057 (0.061)

Table 9: Heckman Models by Region, Summer OG 2000-2016

	Model 24 (Europe)	Model 25 (Asia)	Model 26 (Americas)
Lambda	0.968*** (0.057)	-0.416 (0.454)	4.121*** (1.488)
Selection equation. Dependent variable: Dummy for medals			
Intercept	-3.200*** (0.451)	-1.530*** (0.268)	-2.096*** (0.285)
Exported products	0.018*** (0.002)	0.008*** (0.001)	0.011*** (0.002)
Model characteristics			
Total observations	196	231	193
Censored observations	42	125	125
Year dummies	Yes	Yes	Yes

Notes: Own calculations. Standard errors in parentheses. Year dummies not reported. Matrix for Pacific not positive definite. Convergence criterion was not met for Africa. For Americas MLE model not possible (matrix not positive definite), two-step Heckit used instead.
* Significant at 10 %. ** Significant at 5 %. *** Significant at 1 %.

All in all, despite several differences in tests 14-26, the model appears to be robust and the discrepancies are explainable. This does not mean the approach is without problems. Probably the most well-known drawback of statistical analyses of the Olympic Games is the limit set on the number of national participants in each discipline. Even though there might be numerous world-class 100-meter sprinters in Jamaica, the country is only allowed to nominate three as a maximum. In whitewater slalom the limit is set at one participant per country per event. These participation limits allow relatively lower-performance athletes from other countries to take part in the Games and possibly also to win medals. In general, it could be argued that the participation limits favor smaller countries and this has to be taken into account when trying to forecast medal counts for future Olympic Games. A complex model should therefore include participation restrictions per event; this is well beyond our capacities and is an idea for further research.

6. Conclusions

Sporting success has been known to increase national pride (Denham, 2010), promote national identity (Hong, 2011), lead to higher reputation abroad (Wicker et al., 2012), raise the demand for amateur sport participation (Mutter & Pawlowski, 2014) and provide role models for young people (Lines, 2001). Being one of the most widely watched sporting events in the world, success at Olympic Games can potentially deliver many more positive externalities. However, Olympic success is a multidimensional category and is not only influenced by direct sport policies, but also by numerous exogenous factors. The current paper has examined a wide array of these factors represented by over 140 independent variables and has come to several conclusions:

- The more diversified export structure a country has the higher is its chance to win a medal.
- GDP per capita and the size of economically active population are the most important determinants of Olympic success, robust to virtually any econometric specification of models.
- The role of political factors is ambiguous. While in the Cold War era authoritarian regimes led to better results in sports than democracies, we were not able to persuasively prove this relationship for the period 2000-2016.
- Hosting nations have a certain advantage over other participants. They tend to gain more medals while preparing for their Olympics, at their Olympics as well as four years after their Olympics. The host effect is stronger than the ex-host effect which in turn is stronger than the future-host effect. All these effects seem to be of a bigger importance for large countries than for their smaller counterparts.
- Warmer climate has a negative impact on Olympic performance.

Taking all this into account, what should governments do if they wish to increase the number of medals won at Summer Olympic Games? It appears that direct support of sports should be accompanied by economic measures leading to growth and employment. Open trade policy could be an example of a strategy translating not only into higher exports, but also into higher economic growth and more medals. The positive role of authoritarian regimes cannot be confirmed; our data even appears to show that once a certain level of political freedom is reached, further increases in the level of freedom might lead to better Olympic performance, presumably by fostering competition; however, this very preliminary conclusion should be further tested. Finally, organizing Olympics is another option of increasing a nation's Olympic harvest. For many countries who cannot afford it, perhaps organizing other important international events in Olympic sports might partially bring similar effects.

These results are not absolute. As always, they are subject to limitations of data and statistical approach chosen. While we have explained in detail all the decisions we took, other paths might have been followed as well. Moreover, there is a risk of misspecifying selection and main equations, even though we took all the steps necessary to minimize this possibility.

One idea for further research is to make the temperature variable event-related, i.e. to calculate temperature differences from the host's average annual temperature for each Olympics. The same could be done with latitude, precipitation and other climate factors. Dozens of alternative political variables might also be used. Finally, traditional regression methods could be substituted by rigorous modelling, perhaps on a sub-national or micro level in line with the current trend in some other fields of economics.

References

Andreff, W. (2001). The Correlation between Economic Underdevelopment and Sport. *European Sport Management Quarterly*, *1*(4), 251-279.

---------- (2013). Economic Development as Major Determinant of Olympic Medal Wins: Predicting Performances of Russian and Chinese Teams at Sochi Games. *International Journal of Economic Policy in Emerging Economies*, *6*(4), 314–340.

Ball, D. W. (1972). Olympic Games Competition: Structural Correlates of National Success. *International Journal of Comparative Sociology, 13*(3), 186–200.

BBC. (2016). *Medal Table Rio 2016.* Retrieved from http://www.bbc.com/sport/olympics/rio-2016/medals/countries.

---------- (2012). *Medal Table London 2012.* Retrieved from http://www.bbc.co.uk/sport/olympics/2012/medals/countries.

Berdahl, J. L., Uhlmann, E. L., & Bai, F. (2015). Win–Win: Female and Male Athletes from More Gender Equal Nations Perform Better in International Sports Competitions. *Journal of Experimental Social Psychology, 56*(1), 1–3.

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

Bredtmann, J., Crede, C. J., & Otten, S. (2016). Olympic Medals: Does the Past Predict the Future? *Significance*, *13*(3), 22–25.

Burger, M. J., Van Oort, F., & Linders, G. M. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis*, *4*(2), 167–190.

CEPII. (2011). *GeoDist*. Retrieved from http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=6.

Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). Predicting the Success of Nations at the Summer Olympics Using Neural Networks. *Computers & Operations Research*, *26*(13), 1243–1265.

Database Olympics. (2011). *Database Olympics.* Retrieved from http://www.databaseolympics.com.

Denham, B. E. (2010). Correlates of Pride in the Performance Success of United States Athletes Competing on an International Stage. *International Review for the Sociology of Sport*, *45*(4), 457–473.

Forrest, D., Sanz, I., & Tena, J. D. (2010). Forecasting national team medal totals at the Summer Olympic Games. *International Journal of Forecasting*, *26*(3), 576–588.

Freedom House. (2017). *Freedom in the World 2017*. Retrieved from https://freedomhouse.org/report/freedom-world/freedom-world-2017.

Grimes, A. R., Kelly, W. J., & Rubin, P. H. (1974). A Socioeconomic Model of National Olympic Performance. *Social Science Quarterly, 55*(4), 777–783.

Head, K., & Mayer, T. (2014). Gravity Equations: Workhorse, Toolkit, and Cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.). *Handbook of International Economics*. Amsterdam: Elsevier.

Hoffmann, R., Chew Ging, L. & Ramasamy, B. (2002). Public Policy and Olympic Success. *Applied Economics Letters*, *9*(8), 545–548.

Hong, E. (2011). Elite Sport and Nation-Building in South Korea: South Korea as the Dark Horse in Global Elite Sport. *The International Journal of the History of Sport*, *28*(7), 977–989.

Houlihan, B., & Zheng, J. (2013). The Olympics and Elite Sport Policy: Where Will It All Ends? *The International Journal of the History of Sport*, *30*(4), 338–355.

Jiang, M., & Xu, L. C. (2005). Medals in Transition: Explaining Medal Performance and Inequality of Chinese Provinces. *Journal of Comparative Economics*, *33*(1), 158-172.

Johnson, D. K. N. & Ali, A. (2004). A Tale of Two Seasons: Participation and Medal Counts at the Summer and Winter Olympic Games. *Social Science Quarterly*, *85*(4), 974–993.

Kuper, G. H., & Sterken, E. (2001). Olympic Participation and Performance Since 1896. *SSRN Paper*, Retrieved from https://papers.ssrn.com/sol3/apers.cfm?abstract_id=274295.

Kuppens, T. & Pollet, T.V. (2015). Gender Equality Probably Does Not Affect Performance at the Olympic Games: A Comment on Berdahl, Uhlmann, and Bai (2015). *Journal of Experimental Social Psychology*, *61*(1), 144–147.

Levine, N. (1974). Why Do Countries Win Olympic Medals? Some Structural Correlates of Olympic Games Success: 1972. *Sociology and Social Research*, *58*(4), 353–360.

Lines, G. (2001). Villains, Fools or Heroes? Sports Stars as Role Models for Young People. *Leisure Studies, 20*(4), 285–303.

Lowen, A., Deaner, R. O., & Schmitt, E. (2016). Guys and Gals Going for Gold: The Role of Women's Empowerment in Olympic Success. *Journal of Sports Economics*, *17*(3), 260–285.

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

Mason, C. H., & Perreault, Jr., W. D. (1991). Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research*, *28*(3), 268–280.

Mutter, F., & Pawlowski, T. (2014). Role Models in Sports – Can Success in Professional Sports Increase the Demand for Amateur Sport Participation? *Sport Management Review*, *17*(3), 324–336.

Pfau, W. D. (2006). Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometrics Approach. *Korean Economic Review*, *22*(2), 233–247.

Shughart II, W. F., & Tollison, R. D. (1993). Going for the Gold: Property Rights and Athletic Effort in Transitional Economies. *Kyklos*, *46*(2), 263–272.

The Fund for Peace. (2017). *Fragile States Index 2016*. Retrieved from http://fsi.fundforpeace.org/.

UNCTAD. (2017). *UNCTADstat*. Retrieved from http://unctadstat.unctad.org.

Vagenas, G., & Vlachokyriakou, E. (2012). Olympic Medals and Demo-Economic Factors: Novel Predictors, the Ex-Host Effect, the Exact Role of Team Size, and the "Population-GDP" Model Revisited. *Sport Management Review*, *15*(2), 211–217.

Van Tuyckom, C., & Jöreskog, K. (2012). Going for Gold! Welfare Characteristics and Olympic Success: An Application of the Structural Equation Approach. *Quality & Quantity*, *46*(1), 189–205.

Wicker, P., Hallmann, K., Breuer, Ch., & Feiler, S. (2012). The Value of Olympic Success and the Intangible Effects of Sport Events – A Contingent Valuation Approach in Germany. *European Sport Management Quarterly*, *12*(4), 337–355.

World Bank. (2017). *World Development Indicators*. Retrieved from http://data.worldbank.org/products/wdi.