5-2014

# Olympic Caliber Countries: How Macroeconomic Factors and Previous Performance Impacted the 2012 London Olympic Games

Adam C. Watkins
*University of Tennessee - Knoxville*, awatki18@utk.edu

# Olympic Caliber Countries: How Macroeconomic Factors and Previous Performance Impacted the 2012 London Olympic Games

By: Adam Watkins

# Introduction

The Olympic Games are the largest and most unique of all international sporting competitions. One unique aspect of the games is the amount of preparation a city must undergo to host the games. Bids to host the 2012 Olympic Games were due by July 15, 2003. This means that some of the competitors were only 5 when this decision was made. The 2012 marked the largest games ever to be held with over 10,000 athletes competing from over 204 National Olympic Committees.

This econometric analysis will address four questions: Is there a causal relationship between a country's macroeconomic endowments or a country's historical government type and the ability of that country to win Olympic medals? If so, which macroeconomic factors or government type tends to be the leading driver of this relationship? Additionally, Is there is a relationship between what I call a country's Olympic Factors ( e.g. previous Olympic performance, number of Olympians a country has, and whether or not the country is hosting the games) and the number of medals a country won at an Olympic games. If so, which Olympic factors tend to be the leading driver of this relationship?

The paper will be split into two parts. Part A will involve all of the countries that participated in the 2012 Summer Olympics and will focus on how many medals were won in 2012. Part B focuses only on only those countries that have won at least one medal since 1996 and will look at all the summer Olympics beginning with 1996. The data set for part A will be smaller, since it is only for the 2012 Olympic Games, than Part B's. However, part A will have 19 variables compared to 8 variables in part B. The data set for part B is a panel involving data from all 5 Olympics since 1996 through 2012 for its 8 variables. Due to the difficult nature in finding 11 of the 19 variables used in part A, I excluded them from part B. I felt that fewer variables were a better alternative to possibly skewed results. The data and economic theory and model sections will be shared by both parts. The econometric analysis section and conclusion will be divided into two parts.

# Data

# Part A Variables

| Variable | Description |
|---|---|
| Country | A country Participating in the 2012 Summer Olympic Games |
| 2011 Population | A country's Population between 15 and 64. (2011) |
| Host Country | Whether the Country Hosted the Games (0 – Not the Host, 1 – Yes the Host) |
| Enrollment | The ratio of children of the official secondary school age who are enrolled in secondary school to the population of the official secondary school age. (2009) |
| Urbanwater | The percentage of the population with reasonable access to an adequate amount of water from an improved source, such as a household connection, public standpipe, borehole, protected well or spring, and rainwater collection. (2010) |
| Ruralwater | The percentage of the population with reasonable access to an adequate amount of water from an improved source, such as a household connection, public standpipe, borehole, protected well or spring, and rainwater collection. (2010) |
| Equality | A score ranging from 0 to 1 that describes the level of gender equality in a country. A score of 1 indicated complete gender equality. (2012) |
| GDP_1980 | Per Capita GDP per Country for 1980. |
| GDP_1996 | Per Capita GDP per Country for 1996. |
| GDP_2008 | Per Capita GDP per Country for 2008. |
| ComorFormerCom | A variable describing whether a country has ever been ruled or still is ruled by a communist regime. (0=No they have not, 1=Yes they have) |
| StillCom | A variable describing whether a country is still under a communist regime. (0=No they are |

| | |
|---|---|
| | not, 1=Yes they are) |
| NSCOM | A variable describing whether a country was part of a communist regime other than the Soviet Union that has thus disbanded. (0=No they are not, 1=Yes they are) |
| TRANS | A variable describing whether a country that was previously communist that is currently undergoing the transformation into a democratic regime. (0=No they are not, 1=Yes they are) |
| CEEC | A variable describing whether a country that was previously communist and has successfully undergone the transformation into a democratic regime. (0=No they are not, 1=Yes they are) |
| Previous_medals | The total number of medals a country has won at the Summer Olympic Games starting with the 1996 Atlanta Games. |
| Participants | The number of Olympians that participated for a country in the 2012 London Games. |
| Medals_2012 | The number of medals won by a country at the 2012 London Olympics. |
| Population_1996 | A country's population in 1996. |

# Part B Variables

| Variable | Description |
|---|---|
| Country | A country that has won an Olympic medal since 1996. |
| Total_Medals_By_Year | The amount of medals a country won in a given year. |
| Running_Medal_Count | The amount of medals a country has won since the 1996 Olympics through the specified year e.g. is this variable were for the 2004 Olympics it would be the total of medals from the 1996, 2000, and 2004 games |
| Number_Of_Olympians | The number of Olympians participating from a given country in a given year. |
| Host Country | Whether the Country Hosted the Games (0 – Not the Host, 1 – Yes the Host) |
| ComorFormerCom | A variable describing whether a country has ever been ruled or still is ruled by a |

| | |
|---|---|
| | communist regime. (0=No they have not, 1=Yes they have) |
| Population | The population of a country in a given year. |
| GDP | The GDP of country in a given year. |

I relied heavily on the World Bank as the source for my macro-economic data. The data I obtained from the World Bank includes 1996, 2000, 2004, 2008, 2011 and 2012 population by country, the percentage of the population between 15 and 64 for 2011, the percentage of students enrolled in secondary school, the percentage of the urban population of a country with access to clean drinking water, the percentage of the rural population with access to clean drinking water, GDP per capita for 1980, 1996, and 2008, and GDP for 1996, 2000, 2004, 2008, 2012. In order to find the population of a country between 15 and 64 for 2011, I multiplied the 2011 population by the percentage of the population between the ages of 15 and 64. One downfall in my data is the lack of observations for the percentage of students enrolled in secondary school, percentage of the urban population of a country with access to clean drinking water, and the percentage of the rural population with access to clean drinking water. I looked for this information in places other the World Bank, but that site had the most complete data set.

I gathered my information about the number of medals a country won at the 1996, 2000, 2004, 2008, and 2012 Olympic Games respectively from the sports data site sports-reference.com. I added the number of wins from the 1996 through 2008 Olympic Games to find the previous medals statistic. I decided to focus my Olympic data beginning with 1996 because this is the first Olympics after the collapse of the Soviet Union. Additionally, in order to simplify the country data I combined Serbia, Montenegro, and Kosovo since they were one country in 1996. Data for the number of participants a country had at the 2012 Summer Games came from http://en.wikipedia.org/wiki/2012_Summer_Olympics. While Wikipedia is not always deemed the most reputable source of information, it was the only complete set of participants broken down by country I could find and agreed with the partial lists I found on various other websites. In every economic model I read, there was an idea that the host country always has some sort of unique factor (possibly stemming from home field advantage or national pride) that causes them to win more medals than models would otherwise predict. I used the host country variable, were Great Britain is a 1 and every other country is a 0 to help explain this difference.

Another difficult statistic to acquire was a gender equality statistic. I decided to go with a report published by the World Economic Forum for my statistic. The 2012 report was written by Ricardo Hausmann of Harvard University, Laura D. Tyson of UC Berkeley, Yasmina Bekhouche of The World Economic Forum, and Saadia Zahidi of The World Economic Forum. The world Economic Forum has been publishing this report since 2006 and bases their scores on economic

participation and opportunity education attainment, health and survival, and political empowerment. Similarly to some of the data from the World Bank, this data suffers from lacking an observation for each country.

My last set of data concerns the specification of communist and former communist countries. The country specifications for the StillCom, NSCOM, TRANS, and CEEC variables were obtained from "Is Hosting the Games Enough to Win" by Wladimir Andreff. A country is a member of the CEEC group if they transformed from a centrally planned economy in 1989 or 1990 to a democratic political regime with a market economy and are now part of the European Union. This description includes Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovenia. The Trans group is similar to CEEC group except they lagged or are still lagging on the transformation to a democratic political regime with a market economy. This description includes Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Mongolia, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan and Vietnam. A country belongs to the NSCOM group if the government has more recently or only partially started the democratic transition. This description includes Albania, Bosnia-Herzegovina, China, Croatia, Laos, Macedonia, Montenegro, and Serbia. The StillComm group consists of the two countries that have made no transition to a market economy, Cuba and North Korea. Additionally, I created a variable entitled ComorFormerCom which describes any country that currently is or has ever been part of a communist regime.

One final data downfall with my model is the lack of a variable describing the amount of money a government spends funding Olympic sports. This ideal variable would include the money used towards kids to participating in Olympic activities, Olympic athletes to train, and on facilities for Olympians to train in. This variable would be very helpful in determining how high of a priority the Olympics are to countries since it would give us a firm number on how much they spend compared to other countries. Since I cannot find data for this variable I will assume that all countries support the Olympics at a level directly proportional to their level of income. Thus I will use GDP per capita as a substitute.

# Economic Theory and Model

This topic has been of interest to many economists and statisticians for a number of years, so there is a wealth of information available on the subject matter. From these papers I have determined several factors I am interested in including as explanatory variables. These include GDP per capita, whether the country previously had a communist government in place, total number of previous medals won in Summer Olympic Games, the number of Olympians sent to a particular Olympic Games, population by country, access to clean drinking water, percentage of youth that attend secondary education, and a gender equality statistic. The bulk of my literature research into the topic has come from five papers: *Men, Money, And Medals: An Econometric Analysis of the Olympic Games* by Lui and Suen, *Who Wins Olympic Games: Economic Resources and Medal Totals* by Bernard and Busse, *Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometrics Approach* by Pfau, *Who Wins Olympic Games: Economic Development and Medal Totals* by Bernard and Busse, and *Is Hosting the Games Enough to Win?* by Andreff. While these are the papers I have read most thoroughly, I have skimmed several more.

In part A I will be using four economic models to explore the relationships in my data. The first model will be an OLS regression using only macro-economic and government data to explain the outcome of the 2012 London Games. The second model will be an OLS regression using only Olympic data to explain the outcome of the 2012 London Games. The third model will be an OLS regression using the most important variables as determined by the previous models. The fourth and final model will be a Tobit regression with calculations of the marginal effects of their standard errors using the same variables from the third model to explain the outcome of the 2012 London Games.

In Part B I will be using two economic models to explore the relationships in my panel data. The first model will be a standard OLS regression looking at macro-economic, government, and Olympic data to explain the outcomes of the Olympic games. The second model will be a Tobit regression with calculations of the marginal effects of their standard errors using the same variables from the third model to explain the outcome of the Olympic Games.

# Econometric Analysis
# Part A

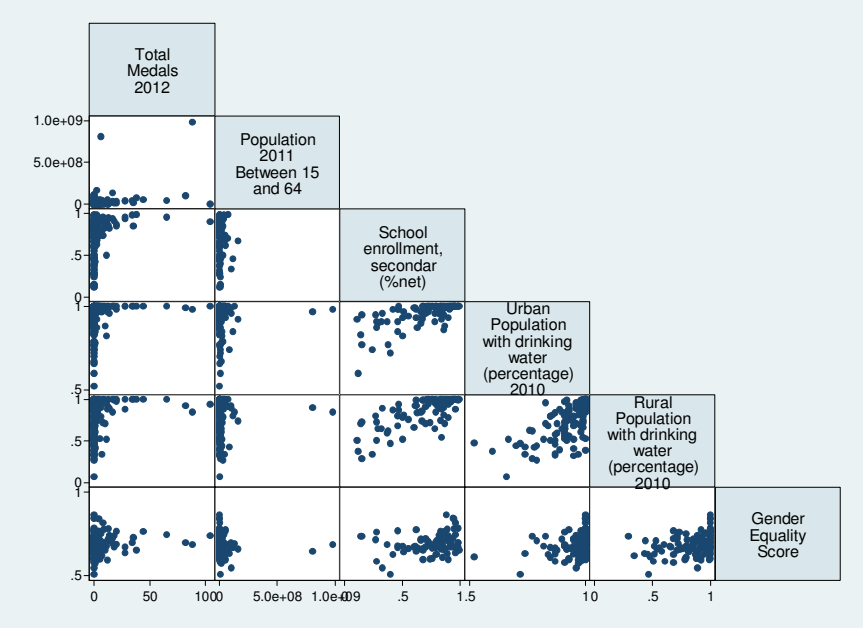


**Figure 1: Linearity of the relationship between Medals_2012 and 5 variables**

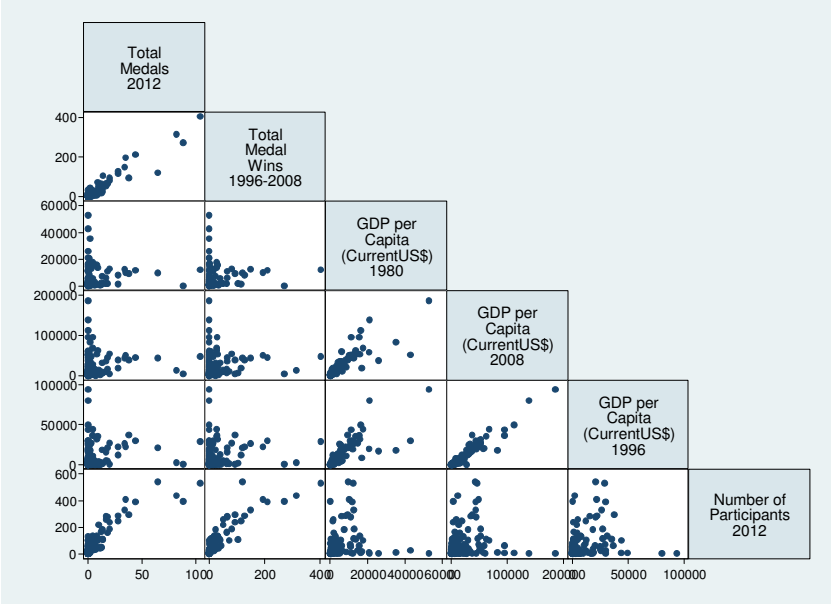


**Figure 2: Linearity of the relationship between Medals_2012 and 5 other variables**

Before I began assembling my models I looked at the linearity relationship between my dependent variable, the number of medals won by a country at the 2012 games, and my

explanatory variables to see if I should consider other functional forms. I came to the conclusion that the ln of the GDP per capita variables should be used in my regressions.

Model 1:
This model is an OLS regression using only macro-economic and government data to explain the outcome of the 2012 London Games. This model will look like Medals_2012 = $ß_0$ + $ß_1$ a population statistic + $ß_2$ enrollment + $ß_3$ urbanwater + $ß_4$ rural water + $ß_5$ equality + $ß_6$ a GDP per capita statistic + $ß_7$ stillcom + $ß_8$ nscom + $ß_9$ nscom + $ß_{10}$ ceec + $E_i$

One important aspect of this model is which population statistic, which GDP per capita statistic, and which government regime statistic to use. As determined above it seemed like the ln of GDP per capita was a better fit for the model. I choose to go with the year 2008 since it best encompassed the Olympian population. Similarly I felt that the population from 1996 best encompassed the Olympic population. I choose to break the communist countries into specific types because this is what most of the papers I looked at had deemed appropriate.

**Table 1: This table provides an OLS regression analysis, which estimates the effects between a change in various independent variable on the number of medals won in the 2012 Olympic Games.**

```
. reg  Medals_2012 population_1996 enrollment urbanwater ruralwater equality  lngdppercapitacurrentus2008  stillcom nscom trans ceec
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 9864.39015 | 10 | 986.439015 | Number of obs | = 78 |
| Residual | 9175.8278 | 67 | 136.952654 | F( 10, 67) | = 7.20 |
| | | | | Prob > F | = 0.0000 |
| Total | 19040.2179 | 77 | 247.275558 | R-squared | = 0.5181 |
| | | | | Adj R-squared | = 0.4462 |
| | | | | Root MSE | = 11.703 |

| Medals_2012 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| population_1996 | 2.16e-07 | 3.02e-08 | 7.16 | 0.000 | 1.56e-07 | 2.76e-07 |
| enrollment | 3.518459 | 14.0824 | 0.25 | 0.803 | -24.59013 | 31.62705 |
| urbanwater | 16.94084 | 43.07199 | 0.39 | 0.695 | -69.0312 | 102.9129 |
| ruralwater | -20.02254 | 16.54584 | -1.21 | 0.230 | -53.04819 | 13.0031 |
| equality | 24.1921 | 22.55848 | 1.07 | 0.287 | -20.83481 | 69.21902 |
| lngdppercapitacurrentus2008 | 4.209118 | 1.949823 | 2.16 | 0.034 | .3172553 | 8.100981 |
| stillcom | 10.80481 | 12.07337 | 0.89 | 0.374 | -13.29374 | 34.90336 |
| nscom | 1.898234 | 8.767922 | 0.22 | 0.829 | -15.60261 | 19.39908 |
| trans | 6.889149 | 5.992205 | 1.15 | 0.254 | -5.071343 | 18.84964 |
| ceec | 1.234087 | 5.707112 | 0.22 | 0.829 | -10.15736 | 12.62553 |
| _cons | -54.60136 | 33.8728 | -1.61 | 0.112 | -122.2118 | 13.00903 |

I chose to use a two-sided 90% level of significance for my analysis. This makes the critical value 1.645. After writing my hypothesis and checking p-values and t-statistics my results are as follows:

$H1_o$: $\beta = 0$; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| Official Variable Name | Descriptive Name | Hypothesis Decision |
|---|---|---|
| Population_1996 | Poplulation in 1996 | Reject Null |
| Enrollment | Enrollment in secondary education | Fail to Reject |
| Urbanwater | Urban population with drinking water | Fail to Reject |
| Ruralwater | Rural Population with drinking water | Fail to Reject |
| Equality | Equality between men and women | Fail to Reject |
| Lngdppercapita2008 | Ln of GDP per capita in 2008 | Reject Null |
| StillCom | Countries that are still communist | Fail to Reject |
| Nscom | Former communist not part of Soviet Union | Fail to Reject |
| Trans | Communist swapping to Democratic | Fail to Reject |
| Ceec | Communist successively become Democratic | Fail to Reject |

Therefore with 90% confidence I reject the null hypothesis that population_1996 and lngdppercapita2008 do not have an effect on the medal outcome of the 2012 London Olympic Games. We can also see that the 95% confidence interval is very good for these two variables. The upper and lower bounds are relatively tight and have the same sign. Therefore we can be very confident that the actual value for these two variables will fall into their respective ranges. Thus, I will use these two explanatory variables in my third and fourth models.

With 90% confidence I fail to reject the null hypothesis that enrollment, urbanwater, ruralwater, equality, stillcom, nscom, trans, and ceec do not have an effect on the medal outcome of the 2012 Summer games. I have decided to exclude enrollment, urbanwater, ruralwater, and equality from models three and four. Even though they were not statistically significant I will use stillcom, nscom, trans, and ceec because literature review and economic theory suggests they should stay.

Model 2:

This model is an OLS regression using only Olympic data to explain the outcome of the 2012 London Games. This model will look like Medals_2012 = $ß_0$ + $ß_1$ previous_medals+ $ß_2$ hostcountry + $E_i$.

**Table 2: This table provides an OLS regression analysis, which estimates the effects between changes in various independent variables on the number of medals won in the 2012 Olympic Games.**

```
. reg  Medals_2012 previous_medals participants hostcountry
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 34807.4641 | 3 | 11602.488 | Number of obs | = 204 |
| Residual | 1902.70746 | 200 | 9.51353729 | F( 3, 200) | = 1219.58 |
| | | | | Prob > F | = 0.0000 |
| Total | 36710.1716 | 203 | 180.838284 | R-squared | = 0.9482 |
| | | | | Adj R-squared | = 0.9474 |
| | | | | Root MSE | = 3.0844 |

| Medals_2012 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| previous_m~s | .2466175 | .0102102 | 24.15 | 0.000 | .2264841 | .2667509 |
| participants | .0028288 | .0057348 | 0.49 | 0.622 | -.0084797 | .0141373 |
| hostcountry | 34.03388 | 3.642045 | 9.34 | 0.000 | 26.85214 | 41.21562 |
| _cons | -.1612038 | .2642793 | -0.61 | 0.543 | -.6823351 | .3599275 |

I chose to use a two-sided 90% level of significance for my analysis. This makes the critical value 1.645. After writing my hypothesis and checking p-values and t-statistics, my results are as follows:

$H1_o$: $\beta = 0$; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| | |
|---|---|
| Previous_Medals | Reject Null |
| Hostcountry | Reject Null |
| Participants | Fail to Reject |

Therefore with 90% confidence I reject the null hypothesis that previous_medals and hostcountry do not have an effect on the medal outcome of the 2012 London Olympic Games. From the regression above we can see that previous medal count has a very small confidence interval, standard error, and standard deviation. This implies this variable is very important in explaining the outcome of the 2012 games. It is highly likely that this will be the driving force for my model. Thus, I will use these two explanatory variables in my third and fourth models.

With 90% confidence I fail to reject the null hypothesis that the number of participants does not have an effect on the null hypothesis. I will not use this variable in models 3 and 4.

Additionally, we can see that the adjusted R-squared is .9474. This is a very high value especially considering there are only three variables involved in this regression. As stated above the participants variable does not have a high level of experience. Therefore we can conclude that the hostcountry variable and previous_medals variables are highly influential in explaining how countries do at the Olympics.

Model 3:

This model is an OLS regression using the important macro-economic, government, and Olympic data as determined by the two previous models to explain the outcome of the 2012 London Games. This model will look like Medals_2012 = $ß_0$ + $ß_1$ lngdppercapitacurrentus2008 + $ß_2$ population_1996 + $ß_3$ previous_medals + $ß_4$ stillcom + $ß_5$ nscom + $ß_6$ nscom + $ß_7$ ceec + $ß_8$ hostcountry + $E_i$.

**Table 3: This table provides an OLS regression analysis, which estimates the effects between changes in various independent variables on the number of medals won in the 2012 Olympic Games.**

```
. reg Medals_2012 lngdppercapitacurrentus2008 population_1996 previous_medals ceec trans nscom stillcom hostcountry
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 35213.171 | 8 | 4401.64638 | Number of obs | = 191 |
| Residual | 1245.46773 | 182 | 6.84322926 | F( 8, 182) | = 643.21 |
| | | | | Prob > F | = 0.0000 |
| | | | | R-squared | = 0.9658 |
| | | | | Adj R-squared | = 0.9643 |
| Total | 36458.6387 | 190 | 191.887572 | Root MSE | = 2.616 |

| Medals_2012 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lngdppercapitacurrentus2008 | .0509067 | .1280878 | 0.40 | 0.692 | -.2018213 | .3036348 |
| population_1996 | 1.31e-08 | 1.93e-09 | 6.79 | 0.000 | 9.30e-09 | 1.69e-08 |
| previous_medals | .2391786 | .0045546 | 52.51 | 0.000 | .2301921 | .2481651 |
| ceec | -2.892379 | .9080613 | -3.19 | 0.002 | -4.68406 | -1.100697 |
| trans | .8421481 | .7738885 | 1.09 | 0.278 | -.6847989 | 2.369095 |
| nscom | 1.731441 | .9740666 | 1.78 | 0.077 | -.1904745 | 3.653356 |
| stillcom | -11.06155 | 2.65791 | -4.16 | 0.000 | -16.30583 | -5.817266 |
| hostcountry | 35.62261 | 2.661927 | 13.38 | 0.000 | 30.3704 | 40.87482 |
| _cons | -.6325078 | 1.094972 | -0.58 | 0.564 | -2.792979 | 1.527963 |

I chose to use a two-sided 90% level of significance for my analysis. This makes the critical value 1.645. After writing my hypothesis and checking p-values and t-statistics my results are as follows:

$H1_o$: β = 0; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| Official Variable Name | Descriptive Name | Hypothesis Decision |
|---|---|---|
| Population_1996 | Poplulation in 1996 | Reject Null |
| Lngdppercapita2008 | Ln of GDP per capita in 2008 | Reject Null |
| StillCom | Countries that are still communist | Reject Null |
| Nscom | Former communist not part of Soviet Union | Reject Null |
| Trans | Communist swapping to Democratic | Fail to Reject |
| Ceec | Communist successively become Democratic | Reject Null |
| Previous-medals | The number of medals won from 1996-2012 | Reject Null |
| Hostcountry | Whether the country was hosting the Olympics | Reject Null |

Therefore with 90% confidence I reject the null hypothesis that previous_medals, hostcountry, stillcom, nscom, ceec, population_1996 and lngdppercapita2008 do not have an effect on the medal outcome of the 2012 London Olympic Games.

With 90% confidence, I fail to reject the null hypothesis that trans does not have an effect on the null hypothesis.

This model shows an $r^2$=.9658 indicating that my model explains 96.58% of the distribution of medals at the 2012 London Olympics. This is a very high R squared value and shows that this model does a very good job of explaining why the 2012 Olympics turned out the way they did. This also means that my model only lacked explaining 3.42% of what affected the outcome of the 2012 Olympic Games.

I will now discuss potential violations of OLS standard assumptions with this model. Since my model does not involve time series or cross-sectional data, I should not be concerned with pure heteroskedasticity or pure serial correlation.

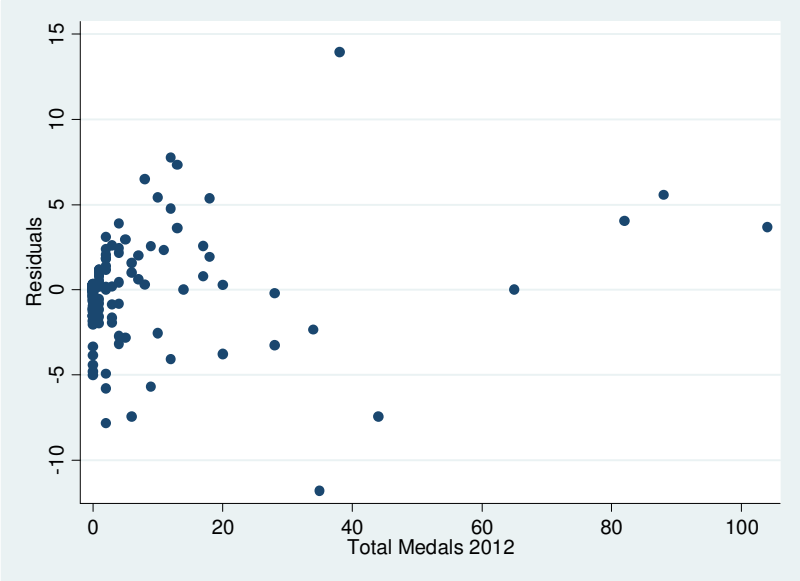


**Figure 3: A graph plotting the residuals from my regression against the total number of Medals won by a country at the 2012 summer Olympic Games.**

There is no distinct pattern in this graph so impure serial correlation should not be a concern.

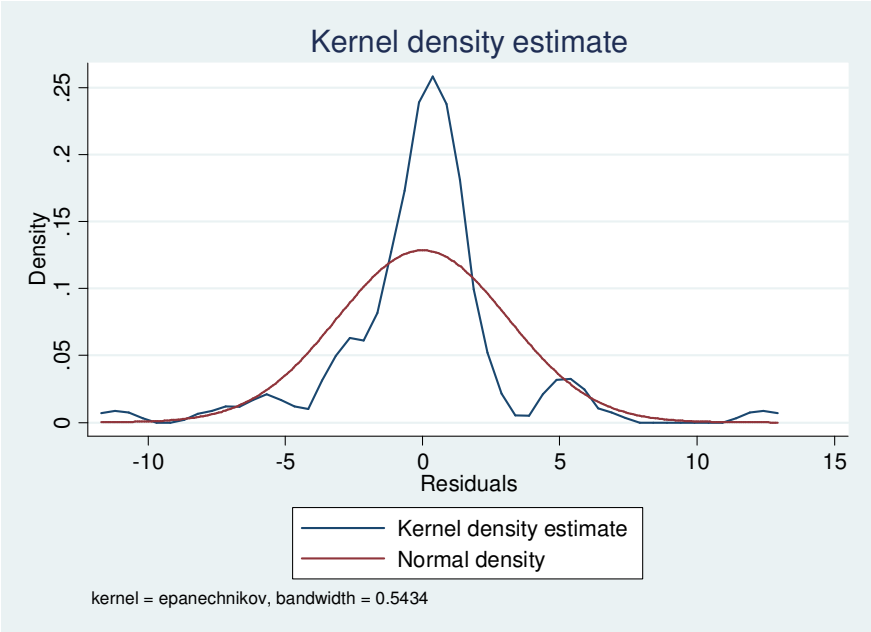


**Figure 4: A graph displaying a normal density curve against the kernel density estimate for my regression.**

This graph shows that there is little concern for impure heteroskedasticity in my data.

**Table 4: This table shows the variance inflation factors for my regression.**

| Variable | VIF | 1/VIF |
|---|---|---|
| previous_m~s | 1.52 | 0.659390 |
| populat~1996 | 1.38 | 0.724506 |
| lngdppe~2008 | 1.19 | 0.842703 |
| nscom | 1.06 | 0.940971 |
| trans | 1.06 | 0.943136 |
| ceec | 1.03 | 0.967721 |
| hostcountry | 1.03 | 0.970842 |
| stillcom | 1.03 | 0.973778 |
| Mean VIF | 1.16 | |

**Table 5: This table shows the correlation matrix for my independent variables.**

```
. correlate lngdppercapitacurrentus2008 population_1996 previous_medals ceec tr
> ans nscom stillcom hostcountry
(obs=191)
```

| | lng~2008 | pop~1996 | previo~s | ceec | trans | nscom | stillcom |
|---|---|---|---|---|---|---|---|
| lngdppe~2008 | 1.0000 | | | | | | |
| populat~1996 | -0.0449 | 1.0000 | | | | | |
| previous_m~s | 0.2817 | 0.4397 | 1.0000 | | | | |
| ceec | 0.1541 | -0.0430 | 0.0540 | 1.0000 | | | |
| trans | -0.1097 | -0.0071 | 0.1236 | -0.0601 | 1.0000 | | |
| nscom | -0.0075 | 0.2311 | 0.0729 | -0.0465 | -0.0565 | 1.0000 | |
| stillcom | 0.0041 | -0.0118 | 0.1220 | -0.0161 | -0.0196 | -0.0152 | 1.0000 |
| hostcountry | 0.0978 | 0.0181 | 0.1433 | -0.0161 | -0.0196 | -0.0152 | -0.0053 |

| | hostco~y |
|---|---|
| hostcountry | 1.0000 |

Since all VIFs are well below 5 and all correlation coefficients are well below .8 and above -.8 multicollinearity should not be a problem in this model.

Model 4:

This model is a Tobit regression using the same important macro-economic, government, and Olympic data from model 3 to explain the outcome of the 2012 London Games. Essentially a Tobit model lets you set a lower bound on the values that can appear in the regression and then will perform a regression on only those values above the stated lower bound. Since it is impossible for a country to win negative medals there is a chance that an OLS model, like the ones used in models 1,2, and 3, will be skewed because the data cuts off at 0. A Tobit model will not have the possibility of incurring this problem because we will set its lower bound at 0. This

model will look like Medals_2012 = $ß_0$ + $ß_1$ lngdppercapitacurrentus2008 + $ß_2$ population_1996 + $ß_3$ previous_medals + $ß_4$ stillcom + $ß_5$ nscom + $ß_6$ nscom + $ß_7$ ceec + $ß_8$ hostcountry + $E_i$.

**Table 6: This chart provides a Tobit regression analysis, which estimates the correlations between the number of medals won at the 2012 London Olympics on the independent variables.**

```
. tobit Medals_2012 lngdppercapitacurrentus2008 population_1996 previous_medals ceec trans nscom stillcom hostcountry, ll(0)

Tobit regression                                Number of obs   =        191
                                                LR chi2(8)      =     315.54
                                                Prob > chi2     =     0.0000
Log likelihood = -262.59788                     Pseudo R2       =     0.3753
```

| Medals_2012 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lngdppercapitacurrentus2008 | .9329296 | .2887285 | 3.23 | 0.001 | .3632648 | 1.502594 |
| population_1996 | 1.68e-08 | 3.20e-09 | 5.24 | 0.000 | 1.04e-08 | 2.31e-08 |
| previous_medals | .2458474 | .0075308 | 32.65 | 0.000 | .230989 | .2607057 |
| ceec | -1.105588 | 1.517742 | -0.73 | 0.467 | -4.100111 | 1.888935 |
| trans | 3.549069 | 1.388997 | 2.56 | 0.011 | .8085612 | 6.289577 |
| nscom | 2.604989 | 1.795019 | 1.45 | 0.148 | -.9366048 | 6.146583 |
| stillcom | -8.255157 | 4.263158 | -1.94 | 0.054 | -16.66642 | .1561037 |
| hostcountry | 36.32255 | 4.25433 | 8.54 | 0.000 | 27.92871 | 44.71639 |
| _cons | -11.75855 | 2.681354 | -4.39 | 0.000 | -17.04889 | -6.468203 |
| /sigma | 4.17667 | .3458095 | | | 3.494383 | 4.858956 |

```
  Obs. summary:        114  left-censored observations at Medals_2012<=0
                        77     uncensored observations
                         0 right-censored observations
```

**Table 7: The table shows the calculations of the marginal effects and their standard errors of the Tobit the number of medals won at the 2012 London Games.**

```
. mfx compute, predict(p(0,.))

Marginal effects after tobit
      y  = Pr(Medals_2012>0) (predict, p(0,.))
         =  .66347485
```

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| lng~2008 | .0815201 | .0257 | 3.17 | 0.002 | .031156 | .131884 | | 8.50169 |
| pop~1996 | 1.46e-09 | .00000 | 5.02 | 0.000 | 8.9e-10 | 2.0e-09 | | 3.0e+07 |
| previo~s | .0214823 | .00138 | 15.61 | 0.000 | .018785 | .024179 | | 18.9058 |
| ceec* | -.1006251 | .14286 | -0.70 | 0.481 | -.380632 | .179382 | | .04712 |
| trans* | .2454812 | .07279 | 3.37 | 0.001 | .102815 | .388147 | | .068063 |
| nscom* | .192138 | .10558 | 1.82 | 0.069 | -.014785 | .399061 | | .041885 |
| stillcom* | -.6059708 | .13737 | -4.41 | 0.000 | -.875203 | -.336738 | | .005236 |
| hostco~y* | .3532972 | .04968 | 7.11 | 0.000 | .255925 | .45067 | | .005236 |

(*) dy/dx is for discrete change of dummy variable from 0 to 1

This table provides the correlation for medal total and each individual variable, holding all other independent variables constant. This allows for an interpretation of the Tobit model similar to the interpretation of an OLS regression.

I chose to use a two-sided 90% level of significance for my analysis. This means that I am looking for a p-value < .1. After writing my hypothesis and checking p-values and t-statistics my results are as follows:

$H1_o$: $\beta = 0$; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| Official Variable Name | Descriptive Name | Hypothesis Decision |
|---|---|---|
| Population_1996 | Poplulation in 1996 | Reject Null |
| Lngdppercapita2008 | Ln of GDP per capita in 2008 | Reject Null |
| StillCom | Countries that are still communist | Reject Null |
| Nscom | Former communist not part of Soviet Union | Reject Null |
| Trans | Communist swapping to Democratic | Reject Null |
| Ceec | Communist successively become Democratic | Fail to Reject |
| Previous-medals | The number of medals won from 1996-2012 | Reject Null |
| Hostcountry | Whether the country was hosting the Olympics | Reject Null |

Therefore with 90% confidence I reject the null hypothesis that previous_medals, hostcountry, stillcom, nscom, trans, population_1996 and lngdppercapita2008 do not have an effect on the medal outcome of the 2012 London Olympic Games.

With 90% confidence I fail to reject the null hypothesis that ceec does not have an effect on the null hypothesis.

While the Tobit regression provided different coefficients in association with the independent variables then the OLS model, the conclusions on which variables are statistically significant remained mostly the same. The only two variables to swap significance were ceec and trans respectively, both of which are indicators of government structure.

# Part B

Part B will address four questions: Is there a causal relationship between a country's macroeconomic endowments or a country's historical government type and the ability of that country to win Olympic medals? If so, which macroeconomic factors or government type tends to be the leading driver of this relationship? Additionally, Is there is a relationship between what I call a country's Olympic Factors ( e.g. previous Olympic performance, number of Olympians a country has, and whether or not the country is hosting the games) and the number of medals a country won at an Olympic games. If so, which Olympic factors tend to be the leading driver of this relationship? Part B is very similar to part A, but with the exception that we will only look at countries that have won at least one medal since 1996. This will test to see if there are different barriers between winning 0 medals to winning at least one medal (Part A) versus winning one medal to winning many medals.

Model 1:

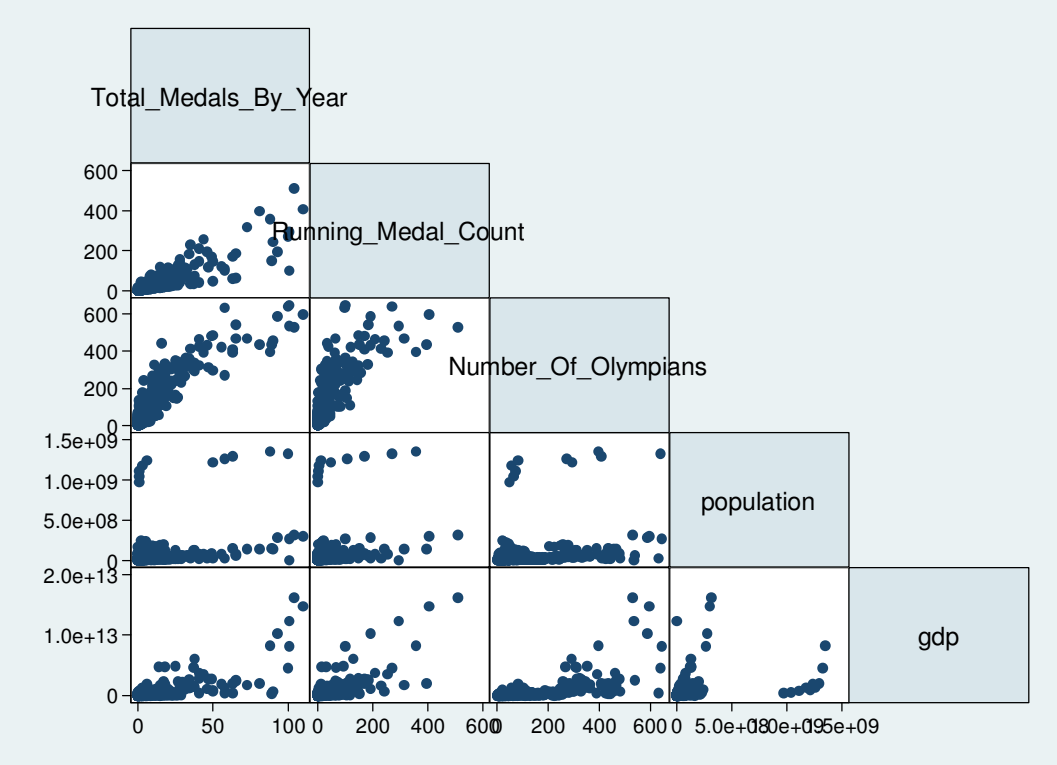


Figure 5: Linearity of the relationship between Total_Medals_By_year and 4 variables

I first looked at the linearity relationship between my dependent variable and the non-dummy explanatory variables. After observing the graphs I see no reason to use any other functional forms in my analysis.

This model is an OLS regression using macro-economic, government, and Olympic data to explain the outcome of the the Olympic games. This model will look like Total_Medals_By_Year

= $ß_0$ + $ß_1$ Running_Medal_Count + $ß_2$ Number_Of_Olympians + $ß_3$ population + $ß_4$ GDP + $ß_5$ comorformercom + $ß_6$ homecountry + $E_i$.

```
. reg   Total_Medals_By_Year Running_Medal_Count Number_Of_Olympians hostcountry comorformercom population gdp
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 129670.058 | 6 | 21611.6763 | Number of obs = | 491 |
| Residual | 15046.0972 | 484 | 31.0869778 | F( 6, 484) = | 695.20 |
| | | | | Prob > F = | 0.0000 |
| | | | | R-squared = | 0.8960 |
| | | | | Adj R-squared = | 0.8947 |
| Total | 144716.155 | 490 | 295.339091 | Root MSE = | 5.5756 |

| Total_Medals_By_Y~r | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Running_Medal_Count | .109702 | .0088204 | 12.44 | 0.000 | .0923709 | .1270331 |
| Number_Of_Olympians | .0669411 | .0035963 | 18.61 | 0.000 | .0598749 | .0740073 |
| hostcountry | 8.514216 | 2.789801 | 3.05 | 0.002 | 3.0326 | 13.99583 |
| comorformercom | 2.803828 | .5989236 | 4.68 | 0.000 | 1.627017 | 3.98064 |
| population | 7.37e-09 | 1.57e-09 | 4.71 | 0.000 | 4.30e-09 | 1.04e-08 |
| gdp | 1.55e-12 | 2.72e-13 | 5.67 | 0.000 | 1.01e-12 | 2.08e-12 |
| _cons | -2.443989 | .3774292 | -6.48 | 0.000 | -3.185591 | -1.702387 |

**Table 8: This table provides an OLS regression analysis, which estimates the effects between changes in various independent variables on the number of medals won in the 2012 Olympic Games.**

I will now discuss potential violations of OLS standard assumptions with this model. The first violation I will test for is heteroskedasticity.

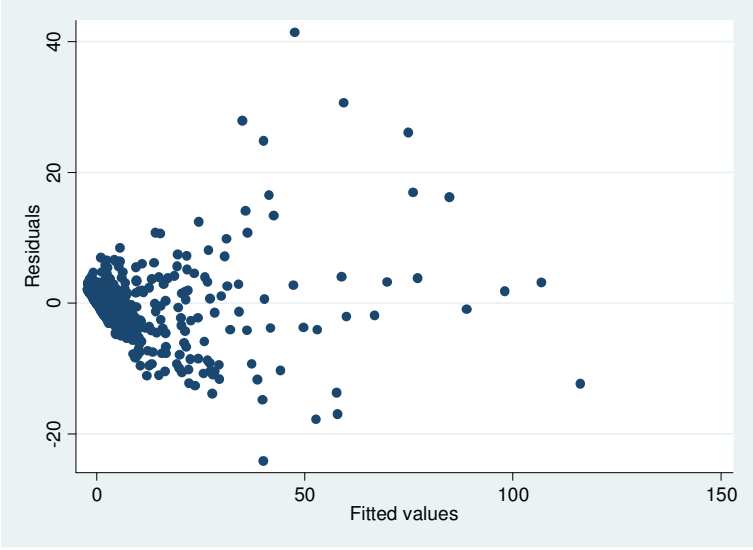


Figure 6: Heteroskedasticity test of plotting residual values versus fitted values.

In order to eliminate this violation I will run the regression again, but with robust standard errors.

```
. reg   Total_Medals_By_Year Running_Medal_Count Number_Of_Olympians hostcountry comorformercom population gdp, robust

Linear regression                                      Number of obs =     491
                                                       F(  5,   484) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.8960
                                                       Root MSE      =  5.5756
```

| Total_Medals_By_Y~r | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Running_Medal_Count | .109702 | .0191506 | 5.73 | 0.000 | .0720735 | .1473305 |
| Number_Of_Olympians | .0669411 | .0069911 | 9.58 | 0.000 | .0532045 | .0806777 |
| hostcountry | 8.514216 | 7.599278 | 1.12 | 0.263 | -6.417434 | 23.44587 |
| comorformercom | 2.803828 | .5842768 | 4.80 | 0.000 | 1.655796 | 3.951861 |
| population | 7.37e-09 | 2.90e-09 | 2.55 | 0.011 | 1.68e-09 | 1.31e-08 |
| gdp | 1.55e-12 | 7.02e-13 | 2.20 | 0.028 | 1.66e-13 | 2.92e-12 |
| _cons | -2.443989 | .4215369 | -5.80 | 0.000 | -3.272257 | -1.615721 |

**Table 9: This table provides an OLS regression analysis using robust standard errors, which estimates the effects between changes in various independent variables on the number of medals won in the 2012 Olympic Games.**

I will now test for impure serial correlation.

```
. plot ress  Total_Medals_By_Year
```

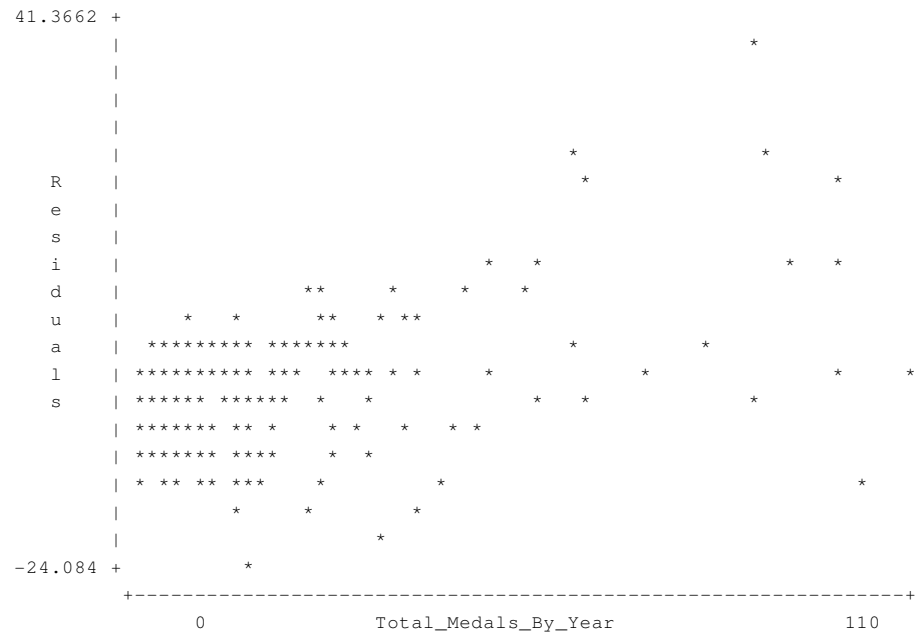


Figure 6: A graph plotting the residuals from my regression against the total number of Medals won by a country at the Olympic Games.

There is no pattern in the graph so impure serial correlation should not be a concern.

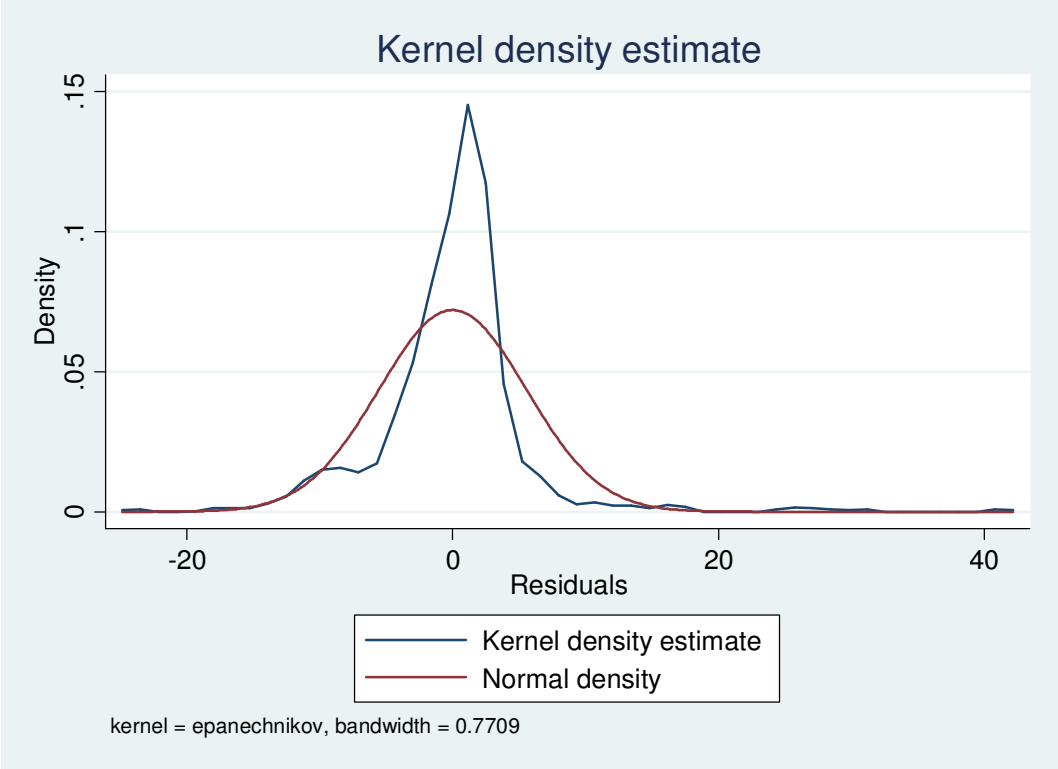


Figure 7: A graph displaying a normal density curve against the kernel density for my regression.

This graph shows there is no concern for impure heteroskedasity in my data.

```
. vif
```

| Variable | VIF | 1/VIF |
|---|---|---|
| Running_Me~t | 3.94 | 0.253922 |
| Number_Of_~s | 3.09 | 0.323110 |
| gdp | 2.57 | 0.389081 |
| hostcountry | 1.24 | 0.807063 |
| population | 1.15 | 0.871060 |
| comorforme~m | 1.09 | 0.915800 |
| Mean VIF | 2.18 | |

Table 10: This table shows the variance inflation factors for my regression.

```
. correlate Running_Medal_Count Number_Of_Olympians hostcountry comorformercom population gdp
(obs=491)
```

| | Runnin~t | Number~s | hostco~y | comorf~m | popula~n | gdp |
|---|---|---|---|---|---|---|
| Running_Me~t | 1.0000 | | | | | |
| Number_Of_~s | 0.7833 | 1.0000 | | | | |
| hostcountry | 0.1988 | 0.3933 | 1.0000 | | | |
| comorforme~m | 0.0893 | 0.0195 | -0.0140 | 1.0000 | | |
| population | 0.3167 | 0.2854 | 0.1668 | 0.0454 | 1.0000 | |
| gdp | 0.7541 | 0.6394 | 0.1832 | -0.1072 | 0.3210 | 1.0000 |

Table 11: This table shows the correlation matrix for my independent variables.

Since all VIFs are well below 5 and all correlation coefficients are between .8 and -.8 multicollinearity should not be problem.

Since we now know that there should be no violations we can summarize the results of the regression. I chose to use a two-sided 90% level of significance for my analysis. This makes the critical value 1.645. After writing my hypothesis and checking p-values and t-statistics, my results are as follows:

$H1_o$: $\beta = 0$; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| | |
|---|---|
| Running_Medal_Count | Reject Null |
| Number_OF_Olympians | Reject Null |
| Hostcountry | Fail to Reject |
| Comorformercom | Reject Null |
| Population | Reject Null |
| Gdp | Reject Null |

Therefore with 90% confidence I reject the null hypothesis Running_Medal_Count, Number_Of_olympians, population, comorformercom, and GDP do not have an effect on the medal outcome of the Olympic Games.

With 90% confidence I fail to reject the null hypothesis that hostcountry does not have an effect on the dependent variable.

Model 2:

This model is a Tobit regression using the same important macro-economic, government, and Olympic data from model 1 to explain the total medals won by each country at each of the Olympic Games. This model will look like Total_Medals_By_Year = $ß_0$ + $ß_1$ Running_Medal_Count + $ß_2$ Number_Of_Olympians + $ß_3$ population + $ß_4$ GDP + $ß_5$ comorformercom + $ß_6$ homecountry + $E_i$.

```
. tobit  Total_Medals_By_Year Running_Medal_Count Number_Of_Olympians hostcountry comorformercom population gdp, ll(0)

Tobit regression                                  Number of obs   =        491
                                                  LR chi2(5)      =    1001.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -1341.7834                       Pseudo R2       =     0.2719
```

| Total_Medals_By_Y~r | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Running_Medal_Count | .1035825 | .0095884 | 10.80 | 0.000 | .0847427 | .1224223 |
| Number_Of_Olympians | .0764545 | .0039892 | 19.17 | 0.000 | .0686164 | .0842927 |
| hostcountry | 5.836239 | 3.032765 | 1.92 | 0.055 | -.12271 | 11.79519 |
| comorformercom | 3.724965 | .6660443 | 5.59 | 0.000 | 2.416283 | 5.033647 |
| population | 7.84e-09 | 1.70e-09 | 4.62 | 0.000 | 4.50e-09 | 1.12e-08 |
| gdp | 1.46e-12 | 2.96e-13 | 4.93 | 0.000 | 8.78e-13 | 2.04e-12 |
| _cons | -4.505232 | .4481488 | -10.05 | 0.000 | -5.38578 | -3.624683 |
| /sigma | 6.041915 | .21213 | | | 5.62511 | 6.45872 |

```
  Obs. summary:         89  left-censored observations at Total_Meda~r<=0
                       402     uncensored observations
                         0 right-censored observations
```

**Table 12: This chart provides a Tobit regression analysis, which estimates the correlations between the number of medals won at the Olympic Games on the independent variables.**

```
Marginal effects after tobit
      y  = Pr(Total_Medals_By_Year>0) (predict, p(0,.))
         =  .91640592
```

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| Runnin~t | .0026346 | .00031 | 8.43 | 0.000 | .002022 | .003247 | | 27.6619 |
| Number~s | .0019446 | .00019 | 10.38 | 0.000 | .001577 | .002312 | | 102.517 |
| hostco~y* | .0754072 | . | . | . | . | . | | .010183 |
| comorf~m* | .0780258 | . | . | . | . | . | | .260692 |
| popula~n | 1.99e-10 | .00000 | 4.34 | 0.000 | 1.1e-10 | 2.9e-10 | | 5.4e+07 |
| gdp | 3.71e-14 | .00000 | 4.63 | 0.000 | 2.1e-14 | 5.3e-14 | | 4.7e+11 |

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

**Table 13: The table shows the calculations of the marginal effects and their standard errors of the Tobit the number of medals won at the Olympic Games.**

This table provides the correlation for medal total in association with X, holding all other independent variables constant. This allows for an interpretation of the Tobit model similar to the interpretation of an OLS regression.

I chose to use a two-sided 90% level of significance for my analysis. This means that I am looking for a p-value < .1. After writing my hypothesis and checking p-values and t-statistics my results are as follows:

$H1_o$: $\beta = 0$; the dependent variables do not have an effect on the number of medals won at the 2012 London Olympics.

$H1_A$: $\beta \neq 0$; the dependent variables do have an effect on the number of medals won at the 2012 London Olympics.

| | |
|---|---|
| Running_Medal_Count | Reject Null |
| Number_OF_Olympians | Reject Null |
| Hostcountry | Fail to Reject |
| Comorformercom | Reject Null |
| Population | Reject Null |
| Gdp | Reject Null |

Therefore with 90% confidence I reject the null hypothesis Running_Medal_Count, Number_Of_olympians, population, comorformercom, and GDP do not have an effect on the medal outcome of the Olympic Games.

With 90% confidence I fail to reject the null hypothesis that hostcountry does not have an effect on the dependent variable.

While the Tobit regression provided different coefficients in association with the independent variables then the OLS model, but the conclusions on which variables are statistically significant remained the same.

# Conclusion

## Part A

This econometric analysis shows that various macro-economic factors, the type of government a country has experienced, previous Olympic experience, and whether or not the country hosted the games explains a significant amount about which countries came out on top and which countries did not at the 2012 London Olympic Games.

My first set of research questions asked: Which macroeconomic factors or government type tends to be the leading driver of Olympic performance at the 2012 London Games? From this analysis it is clear that GDP per capita and populations are important factors in how a country will perform at an Olympic Games. While my regressions showed that lagging GDP per capita by four years provided the best results in explaining 2012 medal performance, the difference was only slight and not enough research was done to back this claim. If further research is done on this topic I believe it would be better to create a variable explain government investment in Olympic sports for kids to participate in, Olympic athletes to train, and on facilities for Olympians to train in. While my model provided the best results when using the population from the birth year required meeting minimum age requirements, more research could be done on the effects of lagging the population variable or using a smaller subset of the population lagged by different time periods. Models 3 and 4 differed slightly on which government regimes where statistically significant to the outcome of the 2012 London games, but it is clear that a country's success at the 2012 Olympic Games was affected to some extent by government history.

My second set of research questions asked: which Olympic factors tend to be the leading driver of Olympic performance at the 2012 Olympic Games? My analysis shows that both a country's Olympic history and whether or not a country is hosting the games are important factors in determining how a county will perform. Previous Olympic performance is the most important addition to my analysis when compared to previous ones. While countries may slowly fall or climb in the total medal rankings, it is rare to see a country make a large change from one Olympics to the next.

# Part B

The econometric analysis in part B looked at the macro-economic factors GDP and population, whether a country has ever been communist, the number of Olypians a country brings, past Olympic performance, and whether or not the country hosted the games to explain the outcome of the Olympic Games since 1996. From the results it seems that a significant amount about which countries came out on top and which countries can be related back to these variables.

Part B eliminates many of the macro-economic factors from part A. The main reason for this is because of the inconsistency between those variables from year to year. Fortunately all of the eliminated variables were found to be insignificant in Part A. After removing these variables, the remaining variables were expanded into a panel set consisting of all the summer Olympics since 1996. From both models in part B it is clear that GDP, population, whether a country has ever been communist, the number of Olympians a country brings, and past Olympic

performance were all important in determining the outcome of the Olympic games. Both models in part B suggest that the host country is not a factor in winning medals at the Olympics. All of these results are consistent with those from part A except the effects of being the host country. I believe that this projection of host country importance is more accurate. The Great Britain, the 2012 host, saw a significant jump in performance in 2008 and 2012. All other hosts from the 1996 through 2012 period had more stable performances throughout all of the years. This and the results from the panel data show that the effects of hosting the Olympics are not that great and the results showing otherwise in part A are because Great Britain is an outlier. I believe this paper has shown that macroeconomic factors do indeed impact performance at the Olympic Games.

References


- Bernard, B. Andrew and Busse, R. Meghan,. (Nov.,2000), Who Wins the Olympic Games: Economic Development and Medal Totals. National Bureau of Economic Research, Working Paper 7998
  < http://www.nber.org/papers/w7998.pdf>

- Bernard, B. Andrew and Busse, R. Meghan,. (Feb.,2004), Who Wins the Olympic Games: Economic Resources and Medal Totals. The Review of Economics and Statistics, Vol. 86, No. 1 pp. 413-417
  < http://www.jstor.org.proxy.lib.utk.edu:90/stable/3211683?seq=2&&>

- Lui, H.-K. and Suen, . (2008), MEN, MONEY, AND MEDALS: AN ECONOMETRIC ANALYSIS OF THE OLYMPIC GAMES. Pacific Economic Review, 13: 1–16. doi: 10.1111/j.1468-0106.2007.00386.x
  < http://onlinelibrary.wiley.com.proxy.lib.utk.edu:90/doi/10.1111/j.1468-0106.2007.00386.x/full>

- Pfau, Wade Donald (2006): Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometrics Approach. Published in: Korean Economic Review, Vol. 22, No. 2 : pp. 233-247.
  < http://mpra.ub.uni-muenchen.de/18829/>

- Wladimir Andreff,. (2012),Is Hosting the Games Enough to Win? A predictive economic model of medal wins at 2014 Winter Olympics. Université Paris1 Panthéon-Sorbonne, Vol. 25, pp. 51-75.
  < http://ideas.repec.org/p/hal/cesptp/halshs-00794057.html>