

Predicting the Olympics Winners

Nicolas Hanss, Alexandre Lauga

ENAC

April 23, 2018

Sommaire

- 1 Introduction
- 2 Forecasting the winners
- 3 What if countries merged?
- 4 Conclusion

Overview

- 1 Introduction
- 2 Forecasting the winners
 - Who would have won the Rio Olympics according to our model?
 - What about Pyeongchang?
 - Limits of our model
 - Better model without outliers ?
- 3 What if countries merged?
 - France and Great Britain
 - Kyrgyzstan-India-Mongolia
- 4 Conclusion

Sample

Sample Collection

n=181 countries

Throughout the Rio and Sochi Olympics.

Sample Variables

Variables

- Population
- HDI
- Mean years of schooling
- Internet penetration
- Unemployment rate
- Income
- Phone

Only for the winter Olympics:

- Minimum monthly temperature
- Maximum altitude

Capturing Human Development

HDI Definition

HDI computed with:

- Life expectancy
- Mean years of schooling
- Gross National Income

Medal count

Medal count

- 9 points for Gold
- 3 points for Silver
- 1 point for Bronze

Sommaire

- 1 Introduction
- 2 Forecasting the winners
 - Who would have won the Rio Olympics according to our model?
 - What about Pyeongchang?
 - Limits of our model
 - Better model without outliers ?
- 3 What if countries merged?
- 4 Conclusion

Analysis of the variables

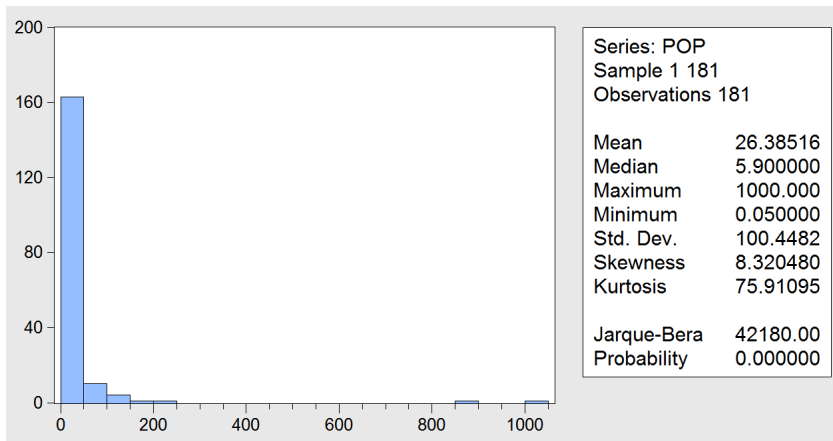
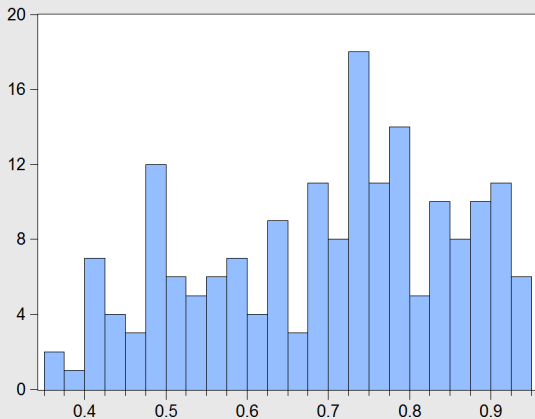


Figure: Population (in millions)

Analysis of the variables

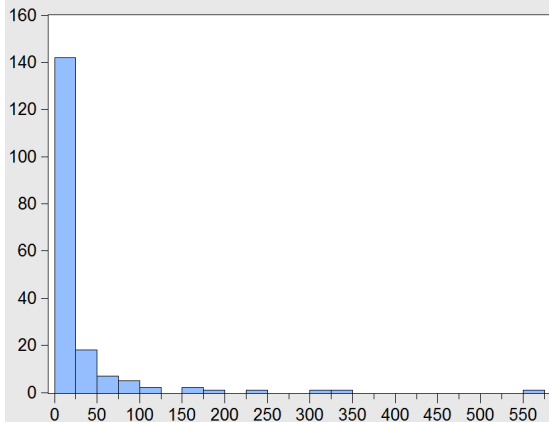


Series: HDI
 Sample 1 181
 Observations 181

Mean	0.697768
Median	0.726000
Maximum	0.949000
Minimum	0.352000
Std. Dev.	0.155324
Skewness	-0.335393
Kurtosis	2.065323

Jarque-Bera	9.981958
Probability	0.006799

Analysis of the variables



Series: RIOSCORE

Sample 1 181

Observations 181

Mean 22.13812

Median 0.000000

Maximum 563.0000

Minimum 0.000000

Std. Dev. 62.37822

Skewness 5.396677

Kurtosis 39.07297

Jarque-Bera 10692.24

Probability 0.000000

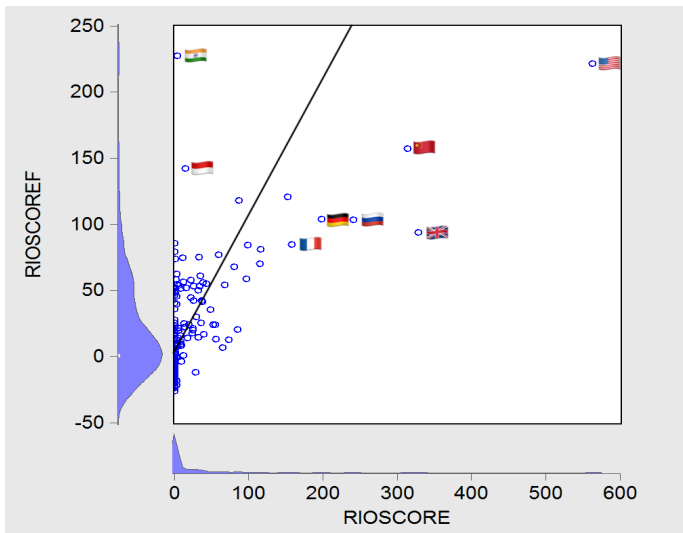
Summer Olympics Explanatory Variables

Explanatory Variables	Expectations	Model
POP	+	+
POP ²	+	-
100*(HDI \geq 0.85)	+	+
SCHOOL	+	+
INTERNET	+	insignificant

Regression model and coefficients

Variables	Model
C	-38.35
POP	0.97***
POP ²	-0.0008***
100*HDI \geq 0.85	0.36***
INTERNET	-0.078
SCHOOL	4.84**
R squared	0.415
F-stat	24.92
Prob.	0.00

Regression Output



Summer Olympics Winners

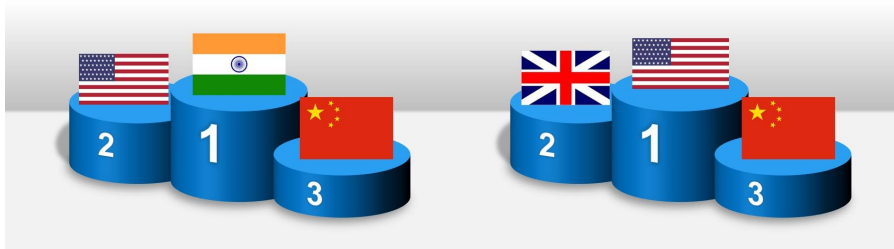


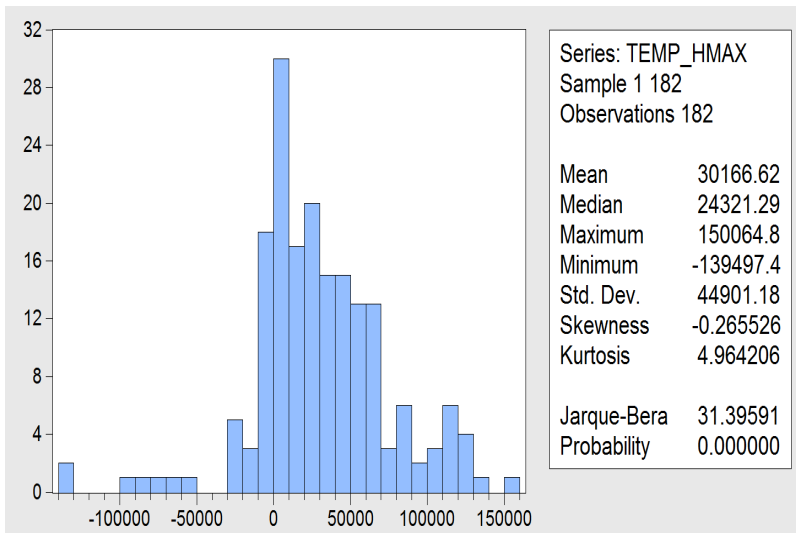
Figure: Expectation

Figure: Reality

What about the winter Olympics?



Analysis of the new variable



Winter Olympics Explanatory Variables

Explanatory Variables	Expectations	Model
POP	+	+++
POP ²	+	-
100*(HDI \geq 0.85)	+++	+
PHONE	+	insignificant
UNEMPLOYMENT	-	insignificant
TEMP	-	insignificant
HMAX	+	insignificant
TEMP*(-HMAX)	+	+

Regression

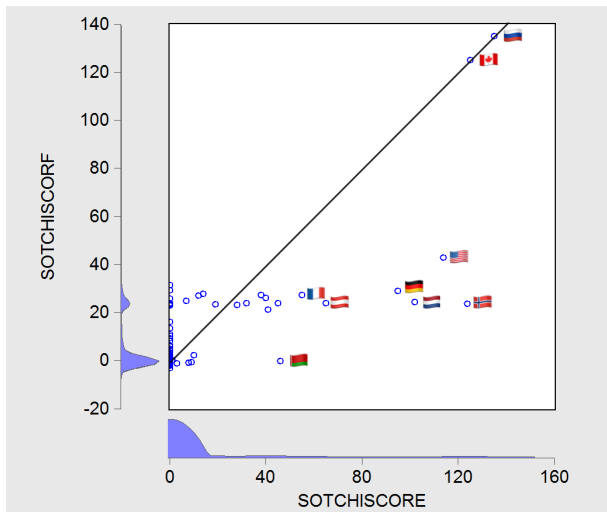




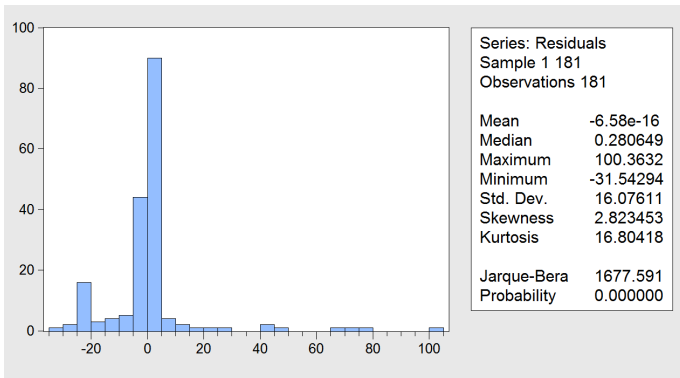
Figure: Expectation

Figure: Reality

Failure of the classical assumptions

Heteroskedasticity Test: White

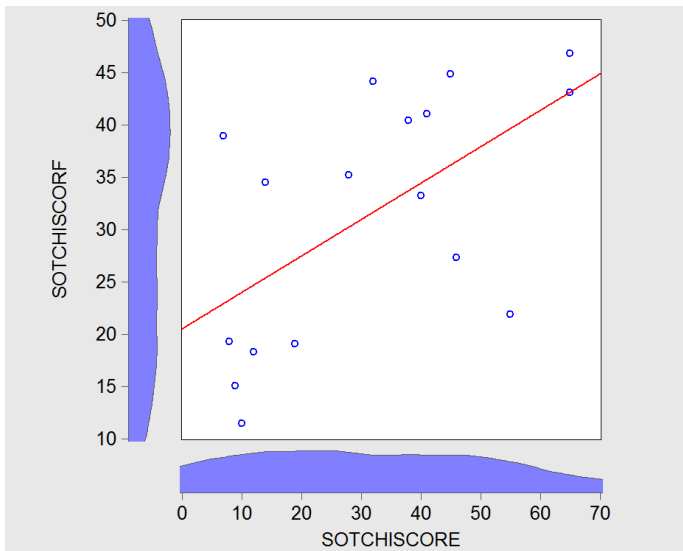
F-statistic	3.706282	Prob. F(21,159)	0.0000
Obs*R-squared	59.48344	Prob. Chi-Square(21)	0.0000
Scaled explained SS	429.4110	Prob. Chi-Square(21)	0.0000



	HDI	INCOME	INTERNET	PHONE	POP	SCHOOL
HDI	1.000000	0.903943	0.913236	0.590272	0.007557	0.906189
INCOME	0.903943	1.000000	0.817998	0.560950	0.016646	0.758104
INTERNET	0.913236	0.817998	1.000000	0.552337	-0.020940	0.825912
PHONE	0.590272	0.560950	0.552337	1.000000	-0.050814	0.484736
POP	0.007557	0.016646	-0.020940	-0.050814	1.000000	-0.032032
SCHOOL	0.906189	0.758104	0.825912	0.484736	-0.032032	1.000000
TEMP	-0.521860	-0.437565	-0.497854	-0.205108	-0.086286	-0.568519
UNEMPLOYME	-0.103032	-0.103768	-0.077618	-0.093497	-0.122368	-0.055946

Figure: Correlation matrix

Without outliers



Sommaire

- 1 Introduction
- 2 Forecasting the winners
- 3 What if countries merged?
 - France and Great Britain
 - Kyrgyzstan-India-Mongolia
- 4 Conclusion

France & Great Britain

France

- POP=40.2
- HDI=0.9
- HMAX=4810
- TEMP=3.9

Great Britain

- POP=41.7
- HDI=0.91
- HMAX=1340
- TEMP=2.8

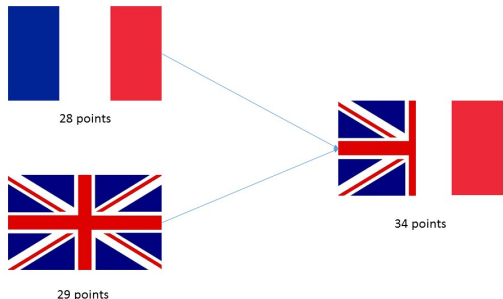


Figure: If France and Great Britain merged

Would merged countries with 0 point score?

Mongolia

- POP=2
- HDI=0.74
- HMAX=4374
- TEMP=-18

India

- POP=860
- HDI=0.62
- HMAX=8850
- TEMP=17.5

Kyrgyzstan

- POP=3.8
- HDI=0.66
- HMAX=7439
- TEMP=-12.7



23 points

Sommaire

- 1 Introduction
- 2 Forecasting the winners
- 3 What if countries merged?
- 4 Conclusion**

Conclusion

Model predictions close to reality

But limits to our model

A few variables preponderant compared to the others

How to improve our model

Look more in details at each discipline

More specific data for each country : measure how they assign importance to different sports

Expand the data with previous Olympics → *time series data*
→ *Take into account the host effect*