

Economic Prediction of Sport Performances: From Beijing Olympics to 2010 FIFA World Cup in South Africa

Madeleine Andreff[†] and Wladimir Andreff^{††}

October 2010

Abstract

This paper uses forecasting techniques to predict outcomes in the Beijing Olympics and 2010 World Cup using economic variables.

JEL Classification Codes: L83

Key Words: Sports, forecasting, Olympics, World Cup

[†] Former Senior Lecturer at the University of Paris-Est Marne la Vallée.

^{††} Professor Emeritus at the University of Paris 1 Panthéon Sorbonne, Honorary President of the International Association of Sport Economists, former President of the French Economic Association (2007-08), andreff@club-internet.fr

ECONOMIC PREDICTION OF SPORT PERFORMANCES: FROM BEIJING OLYMPICS TO 2010 FIFA WORLD CUP IN SOUTH AFRICA

Madeleine Andreff¹ and Wladimir Andreff²

The most usual economic studies about sport mega-events are devoted to their economic impact or cost-benefit outcome. Just like other sport mega-events, FIFA World Cup has been the focus of *ex ante* and *ex post* economic impact studies, and cost-benefit analyses (Ahlert, 2001; Horne, 2004; Kim *et al.*, 2006; Kurscheidt, 2006; Preuss *et al.*, 2008; Rahmann *et al.*, 1998; Szymanski, 2002). It is not surprising that a strand of literature has emerged about what would occur in South Africa in 2010, as far as economic impact is concerned, when hosting the next FIFA World Cup (Bohlmann and van Heerden, 2005; Maennig and du Plessis, 2007). The issue is all the more crucial when it is a developing country that hosts a global sport event (Andreff, 2001 and 2006).

One upstream issue remains unheeded so far with regards to how a nation level of economic development may impact on its sport performance at FIFA World Cup while the same topic about Summer Olympics is well elaborated on in existing economic literature. The latter is based on econometric estimation of how much significant are the economic determinants of the medals won by each participating nation. A same methodology starts developing for other sport mega-events such as international cricket tournaments (Choudhury *et al.*, 2007) and athletics as a specialised component of Summer Olympics (Heyndels and Du Bois, 2006), but not much as regards to FIFA World Cup. Our research question is: would a model based on economic determinants that is used for successfully predicting the distribution of medal wins at Olympic Games be able to provide an econometric prediction of major FIFA World Cup sporting outcomes?

¹ Former Senior Lecturer at the University of Paris-Est Marne la Vallée.

² Professor Emeritus at the University of Paris 1 Panthéon Sorbonne, Honorary President of the International Association of Sport Economists, former President of the French Economic Association (2007-08).

We would respond the aforementioned question as follows. We start with briefly reminding the most interesting methodologies at work in estimating and predicting Summer Olympics medal distribution (1). Then we show how our own model has resolved the issue and predicted how many medals each nation has obtained at the 2008 Beijing Olympics (2). Thereafter, we go on comparing our prediction to actual outcomes of different nations in Beijing and to other prospective econometric studies on the same topic, a comparison which is absolutely rare in the literature so far (3). At this stage, it is necessary to understand why a similar prediction model has not yet developed regarding FIFA World Cup: a major reason is that the soccer World Cup outcome is rather unpredictable, not only due to high uncertainty of outcome, but because each FIFA World Cup final tournament is characterised by a number of “surprises” (4). Then we present how to adapt our model of Olympic medal prediction to FIFA World Cup by introducing, alongside with economic variables compared, some football-specific or “footballistic” variables (5). A model is estimated on the basis of FIFA World Cup results from 1962 to 2006 (6), and then used to provide a prediction of the semi-finalists at the 2010 World Cup in South Africa (7). We conclude with a pinch of salt about the interpretation of such a prediction exercise.

1. Economic determinants of Olympic medals

A widespread assumption across sports economists is that the Olympic performance of a nation must be determined by its endowment in economic and human resources and the development of these resources. Thus, the starting point of most studies about the economic determinants of Olympic medals consists in regressing a nation’s medal wins on its level of GDP per capita and population. Notice that the growth in medal wins by one country logically is an equivalent decrease in medals won by all other nations participating to the Olympics. Therefore, if one wants to understand the Olympic performance of one specific nation, one has to take into account all other participating nations within the overall constraint of the allocated medals total during this year’s Olympics.

A first study of Olympic performance determinants (Jolk *et al.*, 1956) was combining economic variables, such as GDP per capita and population, with weather, nutrition, and mortality in the athlete’s home nation. Later on, in various studies, other variables had been considered as possible determinants of Olympic medal wins: protein consumption, religion, colonial past, newspapers supply, urban population, life expectancy, geographical surface

area, military expenditures, judicial system and those sport disciplines taught at school. However, with the cold war period, another very significant variable emerged: a nation's political regime. The first Western work attempting to explain medal wins by the political regime of nations (Ball, 1972) immediately triggered a Soviet rejoinder (Novikov and Maximenko, 1972), both differentiating capitalist and communist regimes. The first two econometric analyses of Olympic Games (Grimes *et al.*, 1974; Levine, 1974) exhibited that communist countries were outliers in regressing medal wins on GDP per capita and population: they were winning more medals than their level of economic development and population were likely to predict. A last variable has been introduced, namely since Clarke (2000), which is the influence on medal wins of being the Olympics hosting country. The host gains more medals than otherwise expected due to big crowds of national fans, a stronger national athletes' motivation when competing on their home ground and being adapted to local weather, and not tired by a long pre-Games travel.

More sophisticated econometric methodology have been used in most recent studies that predicted Olympic medal wins, such as an ordered Logit (Andreff, 2001), a Probit model (Nevill *et al.*, 2002) or an ordered Probit (Johnson and Ali, 2004); in the latter, a quadratic specification in GDP per capita is employed to capture a postulated inverted U-shaped relationship meaning that higher levels of GDP per capita have a positive but decreasing impact on medal wins below a threshold income level, above which the relationship becomes negative. The most quoted reference is Bernard and Busse (2004) whose Tobit model has been assessed as the most performing one and then used by Jiang and Xu (2005) and Pfau (2006). Bernard and Busse model is considered as the best achieved economic model for estimating and predicting Olympic performance, in which two major independent variables do explain the great bulk of medal distribution across participating countries: GDP per capita and population. Three dummy variables capture a host country effect, the influence of belonging to Soviet-type and other communist (and post-Soviet and post-communist after 1990) countries as against being a non communist market economy. Such dummies are supposed to capture the impact of political regime on medal wins.

2. Predicting Olympic medals distribution in Beijing 2008

Starting from Bernard and Busse, we have elaborated on a more specified model (Andreff *et al.*, 2008) with a few improving emendations. The dependent variable is the number of medal

wins by each nation: $M_{i,t}$ ³. Our first two explanatory variables are GDP per capita in purchasing power parity dollars (PPP \$) and population. Both variables are four year delayed ($t-4$) under the assumption that four years are required to build up, train, prepare and make an Olympic team the most competitive in due time, four years later. That is, for explaining medal wins in 2008, we take the 2004 GDP per capita and population as estimators. A *Host* dummy variable is used to capture the host country effect, *i.e.* the observed surplus of medals usually won by the national squad of the Games hosting nation.

Our first emendation to Bernard and Busse model regards the political regime variable: Bernard and Busse rather crudely divide the world into communist regimes and capitalist market economies which obviously fits with the cold war period. Since then, this is too crude when it comes to the so-called post-communist transition economies (Andreff, 2004 and 2007) in particular with regards to the sports economy sector which has differentiated a lot across former socialist countries during their institutional transformation process (Poupaux and Andreff, 2007). Such differentiation has translated into a scattered efficiency in winning Olympic medals after 1991 (Rathke and Woitek, 2008). Our classification distinguishes first Central Eastern European countries (*CEEC*) which have left a Soviet-type centrally planned economy, in 1989 or 1990, and transformed into a democratic political regime running a market economy: Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia (and Czechoslovakia until the 1993 split), Slovenia, and the GDR (until German reunification in 1990). Another commonality to this group is that these countries have all joined the European Union in 2004 or 2007.

A second country group (*TRANS*) gathers new independent states (former Soviet republics) and some former CMEA member states which have started up a similar process of transition as the CEECs but they are lagging behind in terms of transformation into a democratic regime and some are stalling on the path toward a market economy: Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Mongolia, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan and Vietnam. None of them has joined the EU so far or has really an option to do so. The two next groups have not been Soviet regimes properly speaking in the past, although they have been both communist regimes and planned economies. In the first

³ Bernard and Busse use the percentage of medal wins by each country i for $M_{i,t}$ instead. Our regressions are calculated with both the absolute number of medals (Table 3) and the percentage of medals per country, and the results are not significantly different.

one (*NSCOM*), we sample those countries which have started up a transition process in the 1990s: Albania, Bosnia-Herzegovina, China, Croatia, Laos, Macedonia, Serbia-Montenegro (and the former FSR Yugoslavia before the 1991 split). Two countries have not yet engaged into a democratic transformation and a market economy: Cuba and North Korea and must be considered as still communist regimes (*COM*). All other countries are regarded as capitalist market economies (*CAPME*), the reference group in our estimations. Table 1 exhibits uneven medal distribution according to political regime.

Table 1: Uneven medal distribution by political regime

	1976	1988	1992	1996	2000	2004
<i>CAPME</i>						
Number of medals	268	323	498	543	577	590
Mean (variation coefficient)	3,3 (3,5)	2,2 (4,2)	3,3 (3,7)	3,3 (3,4)	3,5 (3,2)	3,5(3,2)
Number of countries	81	146	151	164	166	168
Countries with $M \geq 1$	35,8%	22,6%	32,5%	33,5%	26,5%	30,4%
<i>NSCOM</i>						
Number of medals	8	40	57	56	59	70
Mean (variation coefficient)	8 (0)	13,3 (1,1)	14,3 (1,9)	8,0 (2,3)	8,4 (2,6)	10 (2,3)
Number of countries	1	3	4	7	7	7
Countries with $M \geq 1$	100,0%	66,7%	50,0%	42,9%	14,3%	42,9%
<i>COM</i>						
Number of medals	15	0	40	30	29	32
Mean (variation coefficient)	7,5 (1,0)	0 (0)	20 (0,8)	15 (0,9)	14,5 (1,4)	16 (1,0)
Number of countries	2	1	2	2	2	2
Countries with $M \geq 1$	100,0%	0,0%	100,0%	100,0%	50,0%	100,0%
<i>CEEC</i>						
Number of medals	195	208	99	91	93	86
Mean (variation coefficient)	32,5 (0,9)	34,7 (1,0)	11,0 (0,9)	9,1 (0,9)	9,3 (0,8)	8,6 (0,7)
Number of countries	6	6	9	10	10	10
Countries with $M \geq 1$	100,0%	100,0%	100,0%	90,0%	90,0%	100,0%
<i>TRANS</i>						
Number of medals	126	132	114	121	143	153
Mean (variation coefficient)	63,0 (1,4)	44,0 (1,7)	38,0 (1,7)	8,6 {(2,0)	10,2 (2,3)	10,9 (2,2)
Number of countries	2	3	3	14	14	14
Countries with $M \geq 1$	100,0%	33,3%	66,7%	71,4%	50,0%	57,1%

In our model, we have introduced a last variable that captures the influence on Olympic performance of a specific sporting culture in a region. For example, Afghan ladies are not used to have much sport participation or to attend sport shows, even less to be enrolled in the Olympic team. As a result of these cultural (sometimes institutional) disparities, some nations are more specialised in one specific sport discipline such as weight-lifting in Bulgaria, Turkey and Armenia, marathon and long distance runs in Ethiopia and Kenya, cycling in Belgium and the Netherlands, table tennis, judo and martial arts in Asia, sprint in Caribbean islands and the U.S., etc. It is not easy to design a variable that would exactly capture such regional sporting culture differences⁴, but we have considered that regional dummies may reflect them. For model estimation, we divide the world into nine “sporting culture” regions: *AFS*, sub-Saharan African countries; *AFN*: North African countries; *NAM*, North American countries; *LSA*, Latin and South American countries; *EAST*, Eastern European countries; *WEU*, Western European countries (taken as the reference region in our estimation); *OCE*, Oceania countries; *MNE*, Middle East countries; and *ASI*, (other) Asian countries.

Table 2: Uneven medal distribution by sporting culture region of the world

	1976	1988	1992	1996	2000	2004
<i>NAM</i>						
Number of medals	122	106	163	156	155	153
Mean (variation coefficient)	12.2 (2.4)	11.8 (2.6)	16.3 (2.1)	15.6 (2.0)	15.5 (1.9)	15.3 (2.1)
Number of countries	10	9	10	10	10	10
Countries with $M \geq 1$	50,0%	30,0%	60,0%	60,0%	60,0%	70,0%
<i>AFN</i>						
Number of medals	0	0	5	6	12	8
Mean (variation coefficient)	0 (0)	0(0)	1 (1,4)	1,2 (1,1)	2,4 (1,0)	1,6 (1,4)
Number of countries	3	5	5	5	5	5
Countries with $M \geq 1$	0,0%	0,0%	40,0%	60,0%	60,0%	40,0%
<i>AFS</i>						
Number of medals	0	10	16	29	42	29
Mean (variation coefficient)	0 (0)	0,3 (5,6)	0,4 (2,4)	0,6 (2,6)	0,9 (2,2)	0,6 (3,0)
Number of countries	3	39	39	45	47	47
Countries with $M \geq 1$	0,0%	5,1%	17,9%	22,2%	21,3%	14,9%
<i>LSA</i>						
Number of medals	5	13	8	23	26	25
Mean (variation coefficient)	0,2 (2,4)	0,5 (2,5)	0,3 (2,4)	0,7 (3,7)	0,8 (3,3)	0,8 (2,7)
Number of countries	23	27	29	32	32	32
Countries with $M \geq 1$	17,4%	25,9%	20,7%	18,8%	12,5%	21,9%
<i>EAST</i>						
Number of medals	328	352	214	217	233	245

⁴ Hoffmann *et al.* (2002a) consider that an important determinant of Olympic successes lies in the degree to which sport and sporting activities are embedded in a nation’s culture. The proxy used to capture such determinant is the total number of times a country has hosted Olympic Summer Games between 1946 and 1998. Our regional variable does not intend to capture only a nation’s sporting culture but how much it is specific (different from the one of nations located in a different geographical area).

Mean (variation coefficient)	41,0 (1,0)	44,0 (1,1)	17,8 (1,7)	8,0 (1,7)	8,6 (2,0)	9,1 (2,0)
Number of countries	8	8	12	27	27	27
Countries with $M \geq 1$	100,0%	100,0%	91,7%	74,1%	55,6%	70,4%
<i>WEU</i>						
Number of medals	110	129	227	243	246	242
Mean (variation coefficient)	5,0 (1,7)	5,4 (1,8)	9,5 (1,9)	10,5 (1,5)	10,3 (1,5)	10,1 (1,4)
Number of countries	22	24	24	24	24	24
Countries with $M \geq 1$	63,6%	58,3%	62,5%	66,7%	58,3%	62,5%
<i>OCE</i>						
Number of medals	9	14	37	47	62	54
Mean (variation coefficient)	2,3 (1,2)	1,3 (3,3)	3,1 (2,6)	3,9 (3,0)	4,8 (3,4)	3,9 (3,4)
Number of countries	4	11	12	12	13	14
Countries with $M \geq 1$	50,0%	9,1%	16,7%	16,7%	15,4%	14,3%
<i>MNE</i>						
Number of medals	2	1	12	11	11	20
Mean (variation coefficient)	0,3 (2,4)	0,1 (3,6)	0,9 (2,0)	0,7 (2,3)	0,7 (2,2)	1,3 (2,2)
Number of countries	6	13	14	15	15	15
Countries with $M \geq 1$	16,7%	7,7%	28,6%	26,7%	20,0%	33,3%
<i>ASI</i>						
Number of medals	36	78	126	109	114	155
Mean (variation coefficient)	2,8 (2,5)	3,4 (2,7)	5,3 (2,4)	4,0 (2,7)	4,4 (2,9)	5,7 (2,5)
Number of countries	13	23	24	27	26	28
Countries with $M \geq 1$	46,2%	26,1%	45,9%	44,4%	19,2%	37,0%

Our first model is simply a specification *à la Bernard and Busse*, but with a differently defined political regime variable. Our estimation is based on a censored Tobit model since a non negligible number of countries that participate to the Olympics do not win any medal. Therefore, a zero value of the $M_{i,t}$ dependent variable does not mean that a country has not participated and we work out a simple Tobit, not a Tobit 2 (with a two stage Heckman procedure). Contrarily to Bernard and Busse, we do not assume that preparing an Olympic team is timeless and, then, independent variables are four years lagged behind the dependent variable. Thus, GDP per capita is noted $(Y/N)_{i,t-4}$, measured in 1995 PPP dollars, and $N_{i,t-4}$ stands for population. Dummies are introduced to test whether the Olympic year is significant, taking 2004 as the reference. These dummies come out to be non significant. In a second model, we adopt a data panel Tobit, in order to take into account unobserved heterogeneity, whose test is significant⁵, and then we opt for estimation with random effects. Our data⁶ encompass all Summer Olympics from 1976 to 2004, except 1980 and 1984 which are skipped out due to boycotts which have distorted the medal distribution by country. Our first specification (1) is:

⁵ A test of maximum likelihood shows that the rho coefficient is significant ($Pr = 0.00$).

⁶ Our data panel is not balanced since the number of existing countries in the world has increased between 1976 and 2004, namely due to the break up of the former Soviet Union, former Yugoslavia and former Czechoslovakia (+ 20 countries), only partly compensated by the re-unification of Germany and Yemen (- 2 countries).

$$M_{i,t}^* = c + \alpha \ln N_{i,t-4} + \beta \ln \left(\frac{Y}{N} \right)_{i,t-4} + \gamma Host_{i,t} + \sum_p \delta_p Political\ Regime_{p,i} + \sum_q \kappa_q Year_{q,i} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim N(0, \sigma^2)$

$$M_{i,t} \text{ observation is defined by } M_{i,t} = \begin{cases} M_{i,t}^* & \text{if } M_{i,t}^* > 0 \\ 0 & \text{if } M_{i,t}^* \leq 0 \end{cases}$$

Our second specification (2) is an emended variant of Bernard and Busse model, including our more specific political regime variable, but also the above described dummies standing for regions of sporting culture ($Region_{r,i}$):

$$M_{i,t}^* = c + \alpha \ln N_{i,t-4} + \beta \ln \left(\frac{Y}{N} \right)_{i,t-4} + \gamma Host_{i,t} + \sum_p \delta_p Political\ Regime_{p,i} + \sum_r \rho_r Regions_{r,i} + u_i + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ and $u_i \sim N(0, \sigma_u^2)$

$$M_{i,t} \text{ observation is defined by } M_{i,t} = \begin{cases} M_{i,t}^* & \text{if } M_{i,t}^* > 0 \\ 0 & \text{if } M_{i,t}^* \leq 0 \end{cases}$$

In a third specification (3), the one used for prediction, we have introduced an additional variable $M_{i,t-4}$ on the right-hand side of model (2), just like Bernard and Busse who do not comment why they proceed in such a way. Our idea is that winning medals at the previous Olympics matters for an Olympic national team which usually expects and attempts to achieve at least as well as four years ago. Such inertial effect is all the more relevant for a nation eager to win as many medals as possible from one Olympiad to the other (a national ‘Olympics cult’⁷) and mobilise a lot of resources to succeed in. The resulting inertia differentiates those nations pulled by Olympics cult from those nations which are used to win zero or few medals. These two groups must be distinguished with using $M_{i,t-4}$ otherwise the prediction will be distorted.

Independent variables	Tobit Model 1	Tobit (panel) Model 2	Tobit Model 3 with lagged M
Log population (t-4)	9,14***	4,15***	2,15***
Log GDP per capita (t-4)	12,42***	5,44***	2,73***

⁷ Which has been fuelled in particular by the cold war, but it has not vanished yet in a number of countries.

Host	24,37***	10,40***	10,04***
Political Regime (ref. CAPME)			
COM	24,34***	11,18***	5,76**
TRANS	23,24***	20,97***	8,15***
CEEC	21,43***	17,94***	6,71**
NSCOM	11,98***	8,06***	5,22*
Region (ref. WEU)			
AFN		-4,45*	-1,81
AFS		3,67*	0,75
NAM		7,93***	0,076
LSA		0,57	-1,08
ASI		-4,34***	-2,58*
EAST		-5,53*	-3,5
MNE		-5,00***	-2,47*
OCE		6,277**	1,3
Year dummy (ref. 2004)			
1976	4,63		
1988	-0,2		
1992	3,33		
1996	3,35		
2000	0,31		
Medals (t-4)			0,95***
Constant	-138***	'-51,30***	-31,57***
Number of observations	941	941	831
Log-likelihood value	-1646,1	-1551,5	-1224,2
Pseudo R2	0,17	0,19	0,34

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

All our estimations deliver significant results (Table 3). In the first Tobit estimation, all coefficients are positive and significant at a 1% threshold, except for year dummies. Thus, it is once again confirmed that medal wins are determined by GDP per capita, population and a host country effect. Political regime is also an explanatory variable, in particular in the case of communist and post-communist transition countries. Our second estimation (Tobit/panel) all in all exhibits the same results. The coefficients of regional sporting culture are significant except for Latin America, an area in which the North American sporting culture may have permeated namely through Caribbean countries and Mexico (classified in *NAM*).

Since Western Europe is the reference a significant coefficient with a positive sign means that a region performs relatively better than Western Europe in terms of Olympic medals (negative sign means lower relative performance than Western Europe). Sub-Saharan Africa, North

America and Oceania perform better. It is a little bit surprising for Sub-Sahara African countries since they are among the least developed in the world (except South Africa), but such effect is due to a few African countries which are extremely specialised in one sport discipline where they are capable to win a non negligible number of medals, such as Ethiopia and Kenya in long distance runs. With negative coefficients, North Africa, Asia, Eastern Europe and Middle East show a lower relative performance than Western Europe. It is not surprising for North Africa and the Middle East due to some restrictions to sporting culture in various countries. In the case of Asia, only few countries are capable to win a significant number of medals (China, both Koreas, Mongolia) given their GDP per capita. A surprise is a negative coefficient of Eastern European countries which are known as outliers or over performers (given their GDP per capita and population). In fact, the negative coefficient results from the variable *Political Regime* which already captures their over performance.

Then, our model (3) is used to predict the medal distribution at Beijing Olympics:

$$M_{i,t}^* = c + \alpha \ln N_{i,t-4} + \beta \ln \left(\frac{Y}{N} \right)_{i,t-4} + \gamma Host_{i,t} + \sum_p \delta_p Political\ Regime_{p,i} + \sum_r \rho_r Regions_{r,i} + \theta M_{i,t-4} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim N(0, \sigma^2)$

$$M_{i,t} \text{ observation is defined by } M_{i,t} = \begin{cases} M_{i,t}^* & \text{if } M_{i,t}^* > 0 \\ 0 & \text{if } M_{i,t}^* \leq 0 \end{cases}$$

Since we use here a pooling estimation⁸ of Model 3, it may suffer from an endogeneity bias and the results may be biased by a correlation between the lagged endogenous variable and the error term. We have treated this issue with a dynamic panel GMM (Arellano and Bond, 1991) which provides robust predictions, and predicted coefficients are close to those estimated with a Tobit. Predictions are published (Andreff *et al.*, 2008) only for a sub-sample of countries⁹ gathered in Table 4.

Table 4 - Prediction of medal wins at Beijing Olympics

	Medals won in 2004	Medal wins predicted in 2008	Lower bound	Upper bound
<i>CEEC:</i>				

⁸ A test of maximum likelihood shows that the rho coefficient is not significant ($Pr = 0.26$) which allows to choose a pooling estimation.

⁹ Result for any other country is available on request addressed to the authors.

Bulgaria	12	12	10	13
Hungary	17	19	17	21
Poland	10	14	12	16
Czech Republic	8	10	8	12
Romania	19	21	19	23
<i>TRANS:</i>				
Belarus	15	17	14	20
Kazakhstan	8	11	8	14
Russia	92	96	93	100
Ukraine	23	27	24	29
<i>NSCOM</i>				
China	63	80	73	86
Cuba	27	29	25	33
<i>CAPME:</i>				
Germany	49	52	50	54
Australia	49	51	47	54
Canada	12	15	13	18
United States	102	106	103	110
France	33	36	35	38
Italy	32	35	34	36
<i>Less developed countries</i>				
Brazil	10	12	10	14
South Korea	30	30	27	32
Kenya	7	2	1	4
Turkey	10	9	7	11

Source: Andreff *et al.*, 2008.

The first-ranked predicted winner is, as usual, the United States, followed by Russia and China, which benefits from a host country effect. Most developed and democratic market economies (CAPME) are predicted to be among the major medal winners together with some post-communist transition countries. Our forecast for France was between 35 and 38 medals while the State Secretary for Sports was hoping that the national team would reach 40.

3. Predictions and actual results: what about the uncertainty of outcome?

The publication of our article in French one month before the opening of Beijing Olympics rapidly became a hit in different French and European media and TV channels. First interviews asked to focus on our model prediction. In a second wave, after the Games end, all interviewers became eager to know for which countries the model had provided a correct or a wrong prediction and, in the latter case, why were it so. This triggered the writing of a follow

up companion paper requested by the French National Institute for Sport and Physical Education (INSEP) to be included in its volume devoted to the overall outcome of Beijing Olympics for France (Andreff, 2009).

Our model provided good predictions regarding those 189 countries for which data were available and computable: 70% of the observed results are in our predicted confidence interval. If one assesses our model prediction as acceptable when its error margin is not bigger than a two medal difference between prevision and reality, then it correctly predicts 88% of all Beijing results. The remaining unexplained 12% (23 nations) account for sporting *uncertainty of outcome*. The model correctly predicts the first ten medal winners, except Japan (instead of Ukraine), misses only four out of the first twenty winners, although with a slightly different ranking. However, the most interesting results are witnessed when the model is clearly wrong in its prediction that is basically for 23 countries, because it means that our five variables (plus the inertial variable) have not captured some core explanation of the Olympics outcome. Fortunately, economists are not capable to predict all the detailed Olympic results, otherwise why still convene the Games?

Which are the major “surprises” delivered by actual results when compared with our predictions? The first one is quite bigger than expected medal wins by the Chinese team – all published predictions have been wrong in this respect. Our model has clearly underestimated the host country effect in China. Possibly, Chinese performance has also been boosted by some undetected doping¹⁰. The second surprise is the underperformance of the Russian Olympic team, the worst since the cold war. It was regarded so much “catastrophic” that Mr. Putin convened the highest decision makers of Russian sport to command a new Olympic policy likely to avoid a repeated disaster at the 2012 London Olympics. In the same vein, some other transition countries, namely Romania, won fewer medals than expected in Beijing. The current state of reforming institutions and restructuring the whole sports sector in these countries (Poupaux and Andreff, 2007) has not been sufficiently captured in our model, despite our more refined political regime variable.

The last three significant surprises are Great Britain, Jamaica and Kenya, the latter being the only two developing countries ranked among the first twenty medal winners. Early

¹⁰ This issue is discussed in depth in Andreff *et al.* (2008) explaining why we had not been able to integrate doping among independent variables despite we wished to do so.

preparation of a super-competitive team for the 2012 London Olympics may have been the cause for higher than expected outcomes of the British team, as it is suggested by Maennig and Wellebrock (2008) who have introduced a “next Olympics host country” variable in their prediction. However, such future host country effect does not improve very much the authors’ forecast: 38 predicted medals as against 47 won by Great Britain. Without such effect our own model predicted between 32 and 35 medals for Great Britain. British medals concentration in cycling (12 medals) may trace back again to undetected doping and/or deep specialisation of a nation in one sport discipline. The latter is the most likely explanation for Jamaican medals concentrated in sprint and Kenyan medals in long distance runs. Though we have taken into account such specialisation through our lagged $M_{i,t-4}$ variable – Kenya had won 7 medals and Jamaica 5 in the same disciplines at Athens Olympics -, the inertia captured with this variable reveals to be insufficient.

Different analyses devoted to Olympics outcomes have been published in 2008. Some are not based on economic approach and econometric forecasting. For instance, Wang and Jiang (2008) have used a strictly mathematical logarithm model, with no economic variable, regarding sport outcomes in one Olympiad as time serial, then data are treated with certain and uncertain time serial models. The prediction is rather good: nothing to say as an economist. Kuper and Sterken (2008) basically present a comparison between different econometric methodologies of medals forecasting and their results. On the other hand, Shibli and Bingham (2008), in a more management science approach, focus on forecasting the number of gold medals that China will win as the host nation. The study is based on an historical reminder of past Chinese Olympic performances concentrated on four sports and the women team, strong government support and the extensive use of performance management principles in preparing the Olympic team. The forecast is restricted to gold medals (46, an underestimation of the 51 actually won by China) and cannot compare with other studies.

We are left with four econometric medal predictions for Beijing Games (Bernard, 2008; Hawksworth, 2008; Johnson and Ali, 2008; Maennig and Wellebrock, 2008) in order to assess the comparative relevance of our own model. We would not compare with Bernard (2008) predictions since they are close to ours, being based on Bernard and Busse modelling as well. All predictions were wrong in two respects: they underestimated actual Chinese medal wins and overestimated medal wins by the Russian and most Central Eastern European national teams. The poorest prediction regarding China (32 medals missing) is Maennig and

Wellebrock's. Their model does not perform well for a number of significant countries: USA (15 medals missing), France (9 medals missing), Kenya (12 medals missing) and Romania (10 medals in excess); it markedly underperforms our forecast for former Soviet Union (namely for Belarus, Kazakhstan and Ukraine¹¹), except Russia, probably because it does not encompass any political regime variable (a weather variable is included instead). However, Maennig and Wellebrock predictions perform not that bad in two respects: they correctly predict exact medal wins for three countries (Finland, Indonesia and Ireland) and exhibit the best percentage of error¹² after our own predictions (see Appendix 1), with a 20.9% error over 19 nations as against a 18.2% error over 21 nations with our model.

Johnson and Ali (2008) dwell upon a model developed by the authors in 2004 in which a differentiation between individual and team Olympic sports is introduced. A last study has been realised by Pricewaterhouse Coopers (Hawksworth, 2008) on behalf of the hosting Chinese Olympic Committee. The econometric methodology is not crystal clearly made explicit but refers to both Bernard and Busse and Johnson and Ali. The variables retained to proceed with the forecasting exercise are: GDP per capita at PPP exchange rates, population, the host nation effect, whether a country was previously part of the former Soviet bloc, and the nation share in medal wins in the previous Olympic Games. These last two studies, regarding their published predictions, are comparable together (and with our model) for 21 nations.

No one of these two predictions had found at least one exact forecast with an equal number of expected and actual medal wins. In comparison, our model correctly predicts 4 results out of 21. Given that these 21 countries have actually won together 648 medals, the percentage of error is 23.1% with Hawksworth forecasting, 34.3% with Johnson and Ali, and 18.2% in our prediction. Bilateral comparison between Hawksworth's published and our expectations can extend to 30 nations. Hawksworth exhibits just one exact country prediction (Cuba) while our model delivers the correct medal wins 9 times out of 30; the error percentage is respectively 25.3% with Hawksworth and 17.6% in our forecast. A comparison between Johnson and Ali (2008) and our results covers 35 nations: no one exact country prediction is found in the

¹¹ Moreover no prediction is published for such a significant country as Cuba.

¹² The percentage of error is calculated as the ratio of cumulative erroneous medal numbers to real medal wins in the sample (or a sub-sample) of participating countries. For instance, with a sub-sample of 21 countries which have won together 648 medals, the denominator is 648. The numerator is 118 erroneously predicted medals with our model (ratio = 18.2%).

former whereas 7 out of 35 predictions are correct in the latter with respectively 43.7% and 23.4% error percentages.

Finally, our model performs pretty well in predicting medal wins at Beijing Olympics when compared to the nearly “official” Pricewaterhouse Coopers forecast on behalf of the Chinese Olympic Committee and to the one with Johnson and Ali model. This confirms that Bernard and Busse methodology is the most performing so far, in particular when slightly emended as regards to the political regime and sporting culture regional variables.

4. Prediction of FIFA World Cup winners: why it is so hard?

The economics of the FIFA World Cup outcome is less developed than the economic approach to Olympic medal wins. There are two ways of explaining international soccer successes in the literature. The most common method is to explain FIFA points and ranking (the FIFA/Coca Cola World Ranking for all national football teams) at one point in time. The second one consists in explaining a nation’s success in FIFA World Cup over time. The economic determinants of the soccer World Cup outcome have only been touched three times in the literature so far (Monks and Husch, 2009; Paul and Mitra, 2008; Torgler, 2004), compared to over thirty papers of this kind devoted to Olympic medal wins.

With the first method, Hoffmann *et al.* (2002b) report regression results identifying the variables influencing a nation’s performance in international soccer, as measured by the January 2001 FIFA points, with a sample of 76 countries that have won medals at the 2000 Sydney Olympics. Inverted U-shape relationships are identified with respect to temperature (climate) and per-capita wealth, and a significant interaction is found between Latin cultural origin and population size, while both variables are individually insignificant. Macmillan and Smith (2007) start with stressing that the non random selection of Hoffmann *et al.*’s sample selects a disproportionate number of countries from the upper end of the international soccer ranking distribution. Thus, they rerun the regression with an augmented sample of 176 countries. They confirm the earlier results with some differences: including a quadratic in population makes it significant as well as the population size. The authors add a history variable that records the year of the first international soccer match in each country, which has a negative and significant effect: there is a penalty to be a late comer in the realm of soccer. An additional dummy for each former Soviet republic except Russia has a negative effect,

which is interpreted as the impact of their integration into the Soviet Union in the early 1920s, thus interrupting their national teams' participation at the international level.

A study by Houston *et al.* (2002), though not devoted to the FIFA World Cup outcome, is of interest regarding one of its explanatory variables. The authors analyse the economic determinants of a countries' leisure and sport proficiency, taking international soccer and 179 countries into consideration. The June 1999 FIFA ranking serves as the dependent variable and is used as a proxy for the proficiency of a given country in international soccer. Independent variables are both economic and 'footballistic': GDP per capita and its square, population, total number of previous World Cup appearances, youth World Cup appearances and the number of years (as a member) in FIFA. The major result is: higher FIFA ranking is associated with higher GDP per capita and bigger population, which converges with our results as regards to the Olympics outcome. The number of World Cup appearances is found to be a positive and significant determinant of leisure proficiency (*i.e.* FIFA ranking).

Yamamura (2009) tests a catching up mechanism between Europe and Latin America, assumed to hold an advanced technology in soccer on the one hand, and developing nations in the football domain on the other hand. The dependent variable is the number of FIFA world ranking points of a nation and data cover the 1993-1998 period. The following explanatory variables are significant: total ranking points in the locality (neighbouring nations) which measure a learning impact from neighbours; average world ranking points for the four major European soccer leagues (England, Germany, Italy, Spain) as a proxy for the most advanced technology level in soccer; logarithm of the years a nation has been FIFA member, real GDP, and population. A catching up effect shows up. The number of World Cup appearances is not significant, contrarily to Houston and Wilson results.

Leeds and Marikova Leeds (2009) dwell upon the previous references and use two dependent variables, FIFA points and the derived FIFA ranking in a cross section regression for 2006 on a sample of 178 countries. They test a number of independent variables but only some emerge as significant. For the purpose of adapting our Olympics model, let us notice that an institutional variable which is longer FIFA membership - assumed to mean that soccer is more embedded in a nation's culture - is not significant while confederation dummies¹³ are not all

¹³ Continental confederations into which FIFA is divided (see below). The authors capture this effect through the impact of placing a club in the confederation championship (like the UEFA Champions League) quarterfinals.

significant. On the other hand, GDP per capita, population, the host country effect, and political regime (communist, former communist, non communist) have a significant impact on international soccer successes, in tune with our model for the Olympics.

Despite the title of his article, ‘The economics of the FIFA football World Cup’, Torgler (2004) does not exactly model the economic determinants of the 2002 soccer World Cup outcome in Japan and South Korea. The dependent variable is a dummy that measures whether a team wins a game or not (1 = win) in the final World Cup tournament. Explanatory variables are not economic. A first variable captures the strength of a team through its FIFA ranking, and the positive influence on success of being the hosting team (home advantage). Since the soccer World Cup final tournaments “are not free of surprises”, a second set of variables is introduced regarding the performance of a team during the game: shots on goal, fouls, corner kicks, free kicks, off sides, cautions, expulsions, actual playing time (based on ball possession). Nothing like GDP per capita or population shows up in the model. The major result is that higher FIFA ranking leads to higher probability of winning the game: a one place improvement in world ranking increases a team’s probability of winning by approximately 1%, but this result is not always statistically significant. Higher number of shots on goal leads to a higher probability of winning; having a referee from the same region has a positive impact on the probability of winning a game, but this effect is not statistically significant¹⁴. From the last result derives the recommendation of not appointing a referee who is from the same region of one of the contending teams.

The first predicting model of the FIFA soccer World Cup outcome actually is the one by Paul and Mitra (2008). It is not much based on economic variables either. The authors start with reminding that in the past four FIFA World Cup tournaments, 1994 to 2006, the top team in FIFA ranking never won, except Brazil in 1994. However, they test the relevance of the last FIFA ranking published before the final round of the World Cup as a benchmark to evaluate teams’ performance. In a first Probit model, the dependent variable is a dummy that measures whether a team wins (1 = win, 0 otherwise) the game or not. The main explanatory variable is FIFA ranking with controlling for the number of goals scored by each team, the number of

¹⁴ We neglect the role of referees in this paper for two reasons: an imperfect referee is a source of competitive unbalance as demonstrated in Groot (2007), and a corrupt referee paves the way for another kind of study about corruption in football. We make the (rather naive) assumption that there is no match fixing and no rigged games even though it is definitely a simplifying assumption in current international soccer (Hill, 2009). Corruption in Italian football has recently attracted the attention of ‘non sport economists’ like Boeri & Severgnini (2009).

yellow cards, and the number of red cards. A second OLS testing considers the scored goal difference as the dependent variable and the rank difference is the main independent variable with controlling for goals scored, the number of yellow and red cards, the number of corner kicks, the number of fouls, the percentage of ball possession, and match attendance. Empirical data are for the 1994 to 2006 World Cups. With regards to the results, higher FIFA ranking is significantly associated with higher probability of winning the game. Higher ranked teams score more goals. A more surprising result is that, though a higher number of yellow or red cards are less likely to win the game, in 2002 and 2006 World Cups countries with more yellow cards were more likely to win the game (and countries with more red cards in the 1998 Cup as well). Other surprises are that more corner kicks and more ball possession are associated with losing the game¹⁵. The authors' conclusion is that overall the favourites or higher ranked teams have the winning trend in their favour, but there is a number of *unexpected match outcomes*, a conclusion joining Torgler's notion of "surprises" and suggesting a rather good competitive balance in FIFA World Cup final tournaments. It is good for the FIFA World Cup uncertainty of outcome but it is why it is so hard to estimate its determinants and make prediction.

The purpose of Monks and Husch (2009) is more to test whether the FIFA World Cup format may lead to a slightly rigged contest or, at least, whether it may favour certain teams, in particular the host country. The paper reminds that in the tournament history, only seven teams have ever won the World Cup (Brazil 5 times, Italy 4, Germany 3, Argentina and Uruguay 2, England and France 1). Of the 18 tournaments held to date, the host has won six times. The authors test the impact of seeding, home continent and hosting on the FIFA Cup outcome from 1982 to 2006. The dependent variable is a national team's World Cup final standing (from the winner down to the 32th among the qualified), and it is regressed on a team's FIFA rank before the World Cup, a dummy variable for being top seeded, a host country dummy, and a dummy variable if the World Cup is being played on a team's own continent. Ex ante rank is positive and significant in determining a team's final standing, as expected. Being top seeded results in an increase in final standing of approximately five places (significant at a 1% level) and the home continent advantage is approximately 2.8 places (but not statistically significant). Both effects probably overlap with the host country variable (the host country is top seeded by definition) which provides three places better than

¹⁵ This may be related with low scoring and defensive tactics on the pitch analysed in Andreff & Raballand (2010); it is to be checked in further research.

the expected final standing, but the result is not statistically significant. Rank, being the host country and playing on one's home continent¹⁶ determine advancement in the tournament to either the quarterfinals or semi-finals.

Now we would take advantage of the previous results to introduce some emendations in our model in view of testing how much the determinants of Olympic medal wins and those of the FIFA World Cup outcome are similar or comparable and, then, attempt at predicting on such basis the outcome of the 2010 World Cup in South Africa.

5. Adapting the Olympics medal model to estimating the determinants of the FIFA World Cup outcome

From the above-mentioned studies it is clear that explaining the FIFA World Cup outcome with socio-economic variables such as GDP, population, institutions or political regime is much harder than explaining Olympics medal wins, for different reasons. Soccer is a sport discipline which is more widespread and developed in some countries (for instance some Latin American countries) than in others, whatever their level of economic development, the size of their population and their democratic or autocratic regime. Such specificity requires the introduction of some 'footballistic' variable in the estimation, contrary to the Olympics which cover so many sport disciplines that overall socio-economic development of a nation affects overall nation outcome, beyond disparities in performance across different sports – thus GDP and population are germane to capture a big share of the determinants. The number of "surprises" is much higher with the soccer World Cup than with the Olympics, first because in one case there is a rather high uncertainty of outcome which pertains to just one sport discipline whereas with the Olympics there are different levels of outcome uncertainty in different sports that, on average, compensate each other for the Olympic teams of big (population) and rich (high GDP per capita) nations.

Another difference between the soccer World Cup and the Olympics lies in the comparative format of the sport contest. In most Olympic disciplines¹⁷, after a preliminary knock-out

¹⁶ All the results are obviously plagued with endogeneity since the final standing is correlated with ex ante ranking (as demonstrated by aforementioned studies) and top seeding is determined by ex ante ranking. Neither an endogeneity test is provided nor a methodology to clean or circumvent it (ex.: endogenous regressor or instrumental variables) is implemented.

¹⁷ Exceptions are team sports and some other sports such as tennis and table tennis.

selection, eight athletes remain in contention for the final and the first three best are allocated (gold, silver and bronze) medals during the final. Thus it is not extremely tricky to build up an estimation of the determinants of medal wins - the first three ranked athletes (nations). It is more complex with the FIFA World Cup final tournament since this contest combines a round robin first stage before the 8th finals and, then, a knock-out second stage from the 8th finals on. The uncertainty of outcome markedly increases from the first to the second stage (Monks and Husch, 2009) and, thus, the impact of socio-economic variables might well dilute a little bit in the course of some knock-out games (thus the “surprises”). This lays ground for the choice of our dependent variable to have it as much comparable as possible with medal wins. We opt for the four nations reaching the semi-finals (*Semifin*) of the soccer World Cup final tournament. In other words, we look for the determinants of being one of the best four ranked teams in the final tournament – and this facilitates using the same estimation model as the one explaining Olympic medal wins. Of course, the best four ranked are the winner, the finalist and two semi-finalists which usually play a ranking game the day before the final¹⁸. Given the dependent variable (being in semi-finals = 1; not being in semi-finals = 0), we estimate a Probit model.

All national teams which have participated to the semi-finals are exhibited in Appendix 2 with their cumulative participation from the first 1930 World Cup in Uruguay up to 2006 in Germany. Retaining the semi-finalists as the dependent variable is also somewhat rooted in FIFA economic incentives. Given the FIFA distribution rules, each team entering the World Cup final tournament earns a 3.79 million € bonus. The next step – reaching the 8th finals – increases this amount by an extra 1.59 million €, followed by an additional 1.90 million € bonus for reaching the quarterfinals. Then for reaching the semi-finals, there is a huge jump of 6.33 million €, followed by only 630,000 € extra to make it for the finals and winning the finals adds another 1.27 million € (Coupé, 2007). Thus, while in sporting terms winning the soccer World Cup is associated with winning the finals, in economic terms, it basically translates into reaching the semi-finals. 8th finalists earn an overall 5.38 million € bonus, quarterfinalists 7.28 million € but semi-finalists earn nearly twice more: 13.61 million € (then 14.64 million € for reaching the final and 15.51 million € for winning it).

¹⁸ In fact, our model can explain the ranking of the soccer World Cup further than the first four best but we focus on the latter as those having participated to semi-finals, finals and ranking matches between the two semi-final losers.

Independent variables are selected with a double purpose in mind: a/ comparing whether the same socio-economic variables play a role in determining the FIFA World Cup outcome as with the Summer Olympics medal wins; b/ finding a sample of socio-economic *and* footballistic variables that explain the soccer World Cup outcome in the long run, in order to come up with an ex post benchmark model that can be used further in ex ante predicting the semi-finalists of the 2010 World Cup. Then, comparing the predicted semi-finalists before June 2010 with the actual ones ex post (after June 2010) will enable detecting the possible “surprises” – diverging results – in a next paper. Due to data availability, the retained observation period runs from the 1962 soccer World Cup in Chile up to the 2006 one in Germany in view of gathering enough data to provide a relevant estimation, which includes 12 FIFA World Cup final tournaments. The data covers all national teams which have participated to soccer World Cup final tournaments since 1962 – that is 16 from the 1962 WC to the 1978 WC in Argentina, 24 countries from the 1982 WC in Spain to the 1994 WC in the USA, and then 32 countries from the 1998 WC in France on, *i.e.* 272 observations in total for each variable in an obviously unbalanced panel.

Population (*Pop*) and GDP per capita (*GDP/cap*)¹⁹ are the first two independent variables considered just like in our Olympics medal model (World Bank data). We add the squares for both variables (Pop^2 and GDP/cap^2), in tune with Houston *et al.* (2002) and Macmillan and Smith (2007), in order to control for possible decreasing returns of population and GDP per capita in terms of soccer World Cup performance. The expectation is that population would have a positive effect on reaching the semi-finals while the specificity of the football discipline may lead to no significant effect of GDP per capita. These variables are introduced in the model with a two year time lag under a similar assumption as with the Olympics: the economic size and level of development of a nation two years ago is the context in which the preparation and training of a national soccer team starts up. Moreover, in the two years after a FIFA World Cup, national teams are used to participate to a regional international contest such as the UEFA Euro or the African Cup of Nations; preparing the World Cup really starts up at the end of such contests (which means in $t-2$), learning from the performance of a national team in such these contests. Other variables of interest are: a World Cup semi-finalists history variable²⁰, FIFA ranking, a host country dummy, a regional variable though

¹⁹ In fact it is more precisely GDP per 1,000 inhabitants that measures this independent variable.

²⁰ Such variable introduces some inertia into the model which is useful when the model is used for prediction.

different from the one in the Olympics medal model, and the number of registered soccer players in the national federations of participating countries²¹.

In previous studies, it has appeared that a nation's history in the football domain, such as World Cup appearances and the length of FIFA membership, matters to explain its international soccer performance. Given our objective of explaining semi-finals participation, we have elaborated on a specific semi-final history variable (*SFstory*) derived from the data in Appendix 2. It is calculated by dividing all the figures in Appendix 2 by the number of FIFA World Cup final tournaments from 1930 up to the year appearing in a column of Appendix 2 (for instance, in the 2006 column, all figures are divided by 18). This variable describes the uneven long-term capacity or competence of a national team to reach the semi-finals in a historical perspective and eventually ranks the nations according to this competence. When one talks about 'footballistic' nations or football-involved countries, Germany, Brazil, or Italy are often mentioned: indeed, they have been the most frequent semi-finalists at FIFA World Cups. As in previous studies, *FIFA rank* is tested as one explanatory variable, taking FIFA ranking one month before the beginning of each World Cup, and a dummy (*Host*) is to capture the impact on performance of being the hosting country.

A regional variable (*Reg*) is different from the one used in the Olympics medal model. The latter's purpose was to capture something like a regional sport culture effect while in the case of FIFA World Cup it must measure the relative strength and density of elite football in six different geographical zones in which the FIFA itself is divided, that is: AFC for the Asian zone, CAF for the African zone, CONMEBOL for the South American zone, OFC for the Oceania zone, UEFA for the European zone, and CONCACAF for the North, Central American, and Caribbean zone. Preliminary qualification to the FIFA World Cup final tournament is organised in different ways (round robin or knock-out or both) in the different zones. Seeding of the final tournament round robin stage varies across years but is based on the successes of teams from each region in previous World Cups and organised in such a way as to assure that top seeded teams will not have to play each other until the second phase (8th finals) of the final tournament (Monks and Husch, 2009). A last assumption to be tested is that a football-oriented nation, that is one in which the number of players is relatively high compared to overall population, must be successful in international soccer. According to a

²¹ Participating countries refer to those participating to the soccer World Cup final tournament. Our model does not attempt to explain the determinants of the qualification to the final tournament so far.

pyramidal explanation of elite sport, stating that the larger the mass of sport participants at the pyramid base, the better the elite top, and most football-oriented nations would have a higher probability to reach FIFA World Cup semi-finals. The number of (registered) soccer players (*Players*) divided by population can capture such possible effect.

Thus, the estimation of the determinants of FIFA World Cup semi-finalists relies on a Probit model:

$$\Pr (Semifin_{i,t}^* = 1) = \Phi [a + b SFstory_{i,t-4} + c N_{i,t-2} + d N_{i,t-2}^2 + e \left(\frac{Y}{N}\right)_{i,t-2} + f \left(\frac{Y}{N}\right)_{i,t-2}^2 + g Host_{i,t} + h FIFArank_{i,t} + \sum_r \rho_r D_r Reg_i + k Players_{i,t}]$$

where Φ is the cumulative normal distribution.

However, the paucity of available data for *FIFArank* and *Players* has led to estimate three different models. FIFA ranking is only available since 1993, when FIFA started calculating and publishing it, whereas the number of registered soccer players in all national federations has been counted and published only in 2000 and 2006 (FIFA Big Count, 2000 and 2006), which markedly reduces the size of the data sample. Thus in a first *M1* model the estimated relationship is:

$$\Pr (Semifin_{i,t}^* = 1) = \Phi [a + b SFstory_{i,t-4} + c N_{i,t-2} + d N_{i,t-2}^2 + e \left(\frac{Y}{N}\right)_{i,t-2} + f \left(\frac{Y}{N}\right)_{i,t-2}^2 + g Host_{i,t} + \rho_1 DEurope_i + \rho_2 DAmSud_i] \quad (1)$$

In a second *M2* model, FIFA ranking is introduced but the sample is reduced to four World Cup final tournaments (1994 to 2006):

$$\Pr (Semifin_{i,t}^* = 1) = \Phi [a + b SFstory_{i,t-4} + c N_{i,t-2} + d N_{i,t-2}^2 + e \left(\frac{Y}{N}\right)_{i,t-2} + f \left(\frac{Y}{N}\right)_{i,t-2}^2 + g Host_{i,t} + \rho_1 DEurope_i + \rho_2 DAmSud_i + l FIFArank_{i,t}] \quad (2)$$

Since FIFA ranking does not show up as statistically significant, it is excluded in a third *M3* model whereas the proportion (percentage) of registered players in the population is taken on board, assuming that this number for 2000 can be used for estimating the FIFA World Cup outcome in 2002:

$$\Pr (Semifin_{i,t}^* = 1) = \Phi [a + b SFstory_{i,t-4} + c N_{i,t-2} + d N_{i,t-2}^2 + e \left(\frac{Y}{N} \right)_{i,t-2} + f \left(\frac{Y}{N} \right)_{i,t-2}^2 + \rho_1 DEurope + \rho_2 DAmSud + m Players_{i,t}] \quad (3)$$

With a small and unbalanced panel, Probit estimation is used as a first step. Then to tackle endogeneity of the semi-final history variable, a Probit model with endogenous regressor (instrumental variables) is resorted to. Valid instruments must be exogenous sources of variation in the semi-finalists, and it is difficult to think of candidate instruments relevant to explain international soccer performance (Macmillan and Smith, 2007). Thus, we retain as instruments those exogenous variables of the best previously estimated model.

6. Socio-economic and ‘footballistic’ determinants of FIFA World Cup semi-finalists

Before estimating *MI* model, a preliminary testing has shown that adding year dummies in *MI* comes out with none of these year dummies being significant. On the other hand, the semi-final history variable is significant at a 1% threshold, population and its square are both significant at 5% with expected signs, and GDP per capita and its square are significant at 10% with expected signs.

Table 5: Estimation of the determinants of the soccer World Cup semi-finalists

Independent variable	Variable name	M1 model	M2 model	M3 model
Semi-final participation history	<i>SFstory (t-4)</i>	1.185 *	2.880 *	4.399 **
Population	<i>Nt-2</i>	0.019 ***	0.004	0.037 **
Population squared	<i>(Nt-2)2</i>	-0.001 ***	-0.00002	-0.0002 *
GDP per capita (1,000 inhabitants)	<i>(Y/N t-2)</i>	0.004	0.012	0.361 *
GDP per capita squared	<i>[(Y/N t-2)]2</i>	-0.0003	-0.001	-0.010 *
Hosting country	<i>Host (t)</i>	1.958 ***	7.089	
Europe region	<i>DEurope</i>	2.233 ***	5.717 ***	0.750
South America region	<i>DAmSud</i>	1.941 ***	4.614 ***	-0.313
FIFA ranking one month before	<i>FIFA rank (t)</i>		-0.013	
Proportion (%) of soccer players	<i>Players (t)</i>			-0.049
Constant		-3.649 ***	-6.175	-5.575 ***
Number of observations		272	120	64
Pseudo R2		0.284 ***	0.361 ***	0.409 ***

*** significant at a 1% threshold; ** at 5%; *at 10%.

Now, estimating *M1* model shows that population and population squared are significant at a 1% threshold; the size of the nation matters. Being the host country is also a significant determinant of reaching the semi-finals. The host country has nearly always advanced past the first round robin group phase of the tournament and rather often qualified for the semi-finals. We have tested the impact of belonging to each of the six regions on to qualifying for the semi-finals; it is not significant for four regions out of six. Taking these four regions as the reference, Europe and South America show up as significant variables at a 1% threshold. Being a European or South American team significantly increases the probability to be semi-finalist (even more in the case of Europe than South America). Most semi-finalists have been either European or South American teams so far. The last significant variable, though only at 10%, is the semi-final history variable. Having participated to past semi-finals has an effect on the probability to reach this stage again. Finally, GDP per capita (per 1,000 inhabitants) and squared GDP per capita are not significant. As expected, this is a major difference between the FIFA World Cup based on a single sport discipline and the multi-sport Summer Olympics. The latter's outcome is influenced by the level of economic development in participating countries whereas the former is not.

In *M2* model, tested from 1994 to 2006, the introduction of FIFA ranking as a supposedly explanatory variable has a devastating effect. Most variables become non significant determinants of being semi-finalist, namely population, population squared and hosting the World Cup. FIFA rank itself is not significant either. The problem with this variable is endogeneity since its calculation includes each team performance (including being semi-finalist) in the last three World Cup²². FIFA ranking also interferes with the semi-final history variable inasmuch as its calculation encompasses the last three semi-finals taken into account in our historical variable. The host country effect fades away from the determinants of qualifying for the semi-finals, against the frequent host nation expectation that its team has a home advantage to qualify for the semi-finals. This must be kept in mind by the Bofana Bofana, *i.e.* the South African soccer team, in particular for a country which does not belong to FIFA European and South American sub-divisions. However, overall, *M2* model is the least satisfying and the most difficult to interpret even though it maintains the European and South American regions as highly significant determinants of reaching the semi-finals. The

²² The calculation formula of FIFA ranking includes, among other, a weighted average of the team's three previous FIFA World Cup results.

semi-final history remains significant at 10% and prevails over FIFA ranking as the relevant footballistic variable to be associated with socio-economic and regional variables in the determination of the FIFA World Cup outcome.

The number of soccer registered players per inhabitant in each participating nation is introduced in *M3* model instead of the FIFA rank. The estimation is run for the last two World Cups, which is in itself a limitation of *M3*. Then, the host variable is automatically dropped because there are only two observations. The number of registered players is not significant which may be interpreted at the world level as follows: soccer mass participation is not a determinant of a nation's participation to the semi-finals of the World Cup final tournament. In other words, this seems to invalidate for soccer the pyramidal view of sport where the larger the pyramid base of mass participation, the higher (the better) the elite top of the pyramid, *i.e.* the performance in international contests. On the other hand, population is significant again, the semi-final history variable is even more significant (at 5%) than in previous models while GDP per capita and squared become significant at 10%. However, the regional variables, Europe and South America, are not significant because only two World Cups are kept: in 2006, no South American team has reached the semi-finals whereas in 2002 one semi finalist was neither European nor South American (South Korea).

Finally, we have to check for endogeneity between the dependent variable and one explanatory variable, the semi-final history. The latter is influenced by each new World Cup results (semi-finals), though in the long run these results have an decreasing marginal effect on our cumulative variable in contrast with FIFA ranking (which is a weighted mobile mean). Thus, we use the semi-final history as an instrumented variable and all other variables taken on board in *M1* model as instruments. First, the semi-final history variable is explained by population, population squared, GDP per capita and squared, hosting the Cup and regional variables (Europe and South America), and then the relationship between the dependent variable (qualifying for the semi-finals) and the endogenous regressor 'semi-final history' is studied.

Table 6 shows that all the instrumental variables are explanatory of the semi-final participation history except the host variable. It is logical since the semi-final history variable is a cumulative percentage over 18 Cups whereas a country has been hosting the Cup only

once - or twice²³. Now the model is quite consistent and close to our Olympics medal model since not only population and regional variables but also GDP per capita are highly significant determinants of the FIFA World Cup outcome. In fact, the only clear specificity of the soccer World Cup is that being the host country is not a comparable advantage to the one of hosting Summer Olympics. However, such reality has been blurred for a long time by the World Cup being always located either in Europe or South America until 1990. Since then, the number of exceptions to this former rule is increasing with a location in North America (1994), Asia (2002) and Africa (2010).

Table 6 – Instrumental variables explaining the semi-final history variable

Semi-final participation history	<i>SFstory (t-4)</i>	Coef.	P > t
Population	<i>Nt-6</i>	0.001	0.001 ***
Population squared	<i>(Nt-6)2</i>	-7.98e-07	0.001 ***
GDP per capita (1,000 inhabitants)	<i>(Y/N t-6)</i>	0.005	0.005 ***
GDP per capita squared	<i>[(Y/N t-6)]2</i>	0.0001	0.024 **
Hosting country	<i>Host (t)</i>	0.079	0.144
Europe region	<i>Europe</i>	0.147	0.000 ***
South America region	<i>AmSud</i>	0.234	0.000 ***
Constant		-0.828	0.000 ***
Number of observations		256	
Wald Chi 2		27.77 ***	

The relationship between qualifying to the semi-finals and the semi-final history

Qualifying to the semi-finals		Coef.	P > t
Semi-final participation history	<i>SFstory (t-4)</i>	5.536	0.000 ***
Constant		-1.611	0.000 ***

*** significant at a 1% threshold; ** at 5%

7. The prediction for the 2010 FIFA World Cup in South Africa

The model estimated with instrumented and instrumental variables as well as *MI* model are now used to build up our prediction of the 2010 FIFA World Cup semi-finalists, taking the data for population and GDP in 2008, the cumulative semi-final history variable up to 2006, the host country (South Africa) and the regional variables (Europe, South America). The predictions are exhibited in Table 7.

²³ The FIFA World Cup final tournament has been hosted twice in France (1938, 1998), Germany (1974, 2006), Italy (1934, 1990) and Mexico (1970, 1986).

Table 7 - Prediction of the four semi-finalists at the 2010 FIFA World Cup

Model with instrumental variables			M1 model			FIFA rank
Teams	Rank	Proba*	Teams	Rank	Proba*	March 2010
Germany	1	96.2%	Germany	1	55.6%	1. Spain
Brazil	2	92.9%	Italy	2	42.2%	2. Brazil
Italy	3	80.2%	Brazil	3	41.4%	3. Netherlands
France	4	47.1%	France	4	35.7%	4. Portugal
Most probable quarterfinalists						
Argentina	5	35.2%	England	5	22.1%	5. Italy
Uruguay	5	35.2%	Argentina	6	21.2%	6. Germany
Netherlands	7	24.6%	Spain	7	18.7%	7. England
England	8	16.0%	South Africa	8	17.5%	8. France
Most probable 8th finalists						
Serbia	8	16.0%	Serbia	9	13.8%	9. Argentina
Portugal	8	16.0%	Portugal	10	12.6%	10. Greece
Spain	11	9.6%	Netherlands	11	10.7%	11. Chile
Chile	11	9.6%	Chile	12	8.8%	12. Serbia
<i>South Korea</i>	11	9.6%	<i>Slovakia</i>	13	8.6%	13. USA
USA	11	9.6%	Uruguay	14	8.4%	14. <i>Mexico</i>
South Africa	15	5.4%	Greece	15	8.3%	15. Uruguay
Greece	15	5.4%	<i>Slovenia</i>	16	6.5%	16. <i>Australia</i>

* Probability to participate to semi-finals.

The first four teams with the highest probabilities to participate to the semi-finals in South Africa are the same with both *M1* model and the model tested with instrumental variables. If one interprets the two highest ranks (probabilities) as the most probable finalists, the former predicts Germany playing Italy in the final while the latter forecasts Germany playing Brazil. France is ranked fourth in both cases. Compared to the last published FIFA ranking (March 2010) these results are strikingly different: the first four FIFA ranked teams are Spain and Brazil (potential finalists), then the Netherlands and Portugal. From this comparison, it seems that Brazil is the most widely admitted semi-finalist whatever the methodology used for prediction.

How interpret the three differences between our models' prediction and FIFA ranking? Our models derive the probabilities and rankings from long run economic, regional and footballistic (semi-final history) variables whereas FIFA ranking is based on most recent soccer performances during the past three years, as of March 2010. The latter is less stable

and may well slightly change as of June 2010 since it depends on the game outcomes of all played games by all national teams meanwhile. Our rankings will be the same in June 2010. However the three rankings might well be at variance with the actual results of the 2010 FIFA World Cup because there is not such a thing as uncertainty of outcome in either our modelling or the FIFA points calculation. Then, in July 2010, it would be time to compare predictions with actual semi-finalists and derive from the possible observed differences which had been the “surprises” in the 2010 soccer World Cup outcome taking our model prediction as a benchmark.

Although it is beyond the core purpose of our prediction, it is not without interest to note that 13 out of 16 national teams are common to the three rankings exhibited in Table 7. If one goes as far as interpreting these rankings as a probability to participate to the 8th finals, there is a good chance that Argentina, Brazil, Chile, England, France, Germany, Greece, Italy, the Netherlands, Portugal, Serbia, Spain, and Uruguay would qualify for the second stage of the 2010 soccer World Cup final tournament. Since our two models encompass a host country effect, both predict South Africa qualifying for the 8th finals contrarily to this nation current FIFA ranking (88th in March 2010). The US team makes it to the 8th finals with our instrumental variable model and FIFA ranking (16th in March 2010). It would be more of a surprise if Slovakia and Slovenia (*MI* model prediction), South Korea (instrumented model), Mexico and Australia (FIFA ranking) were to qualify for the 8th finals in South Africa. Of course, those twelve teams²⁴ which are not mentioned in Table 7 would be “big surprises” if they reached the semi-finals.

Conclusion

Comparing the estimated determinants of Olympics medal winning nations and FIFA semi-finalist nations paves the way for some confidence in basic economic variables (GDP per capita and population) as good predictors of sporting outcomes in both cases. Regional variables, though not identically structured, also matter in the determination of Olympics medal winners and soccer semi-finalists. Hosting the world sport contest is absolutely determinant in Summer Olympics where the host country benefits from a substantial bonus in terms of medal wins (this effect was very strong at the 2008 Beijing Olympics) whereas a

²⁴ That is Algeria, Cameroon, Denmark, Ghana, Honduras, Ivory Coast, Japan, New Zealand, Nigeria, Paraguay, RDP (North) Korea, and Switzerland.

similar impact does not play a significant – or only a much milder – role in the FIFA World Cup final tournament. This reflects in the predictions drawn from our modelling: hosting China was predicted as one of the three big winners at the Beijing Olympics while South Africa is not expected to qualify either for the semi-finals with *MI* model or the quarterfinals with the instrumented model. Though our modelled predictions had been able to correctly detect between 70% and 88% of actual medal winners at the Beijing Games, they should not be interpreted, in particular in the case of the 2010 FIFA World Cup, as a prognostic about which teams will be the actual semi-finalists since no variable captures the uncertainty of outcome which obviously could create divergences – the so-called “surprises” – compared with our benchmarking models. We have surveyed in this paper a number of reasons why it is so hard to predict the outcome of just one single sport discipline’s world contest such as the soccer World Cup. Any economic prediction of sporting performance must be taken with a pinch of salt; with regards to FIFA World Cup, better a pint than a pinch. The uncertainty of outcome seems to be much higher in the soccer World Cup than in Summer Olympics, a preliminary conclusion which deserves further research.

References:

- Ahlert G. (2001), The Economic Effects of the Soccer World Cup in Germany with Regard to Different Financing, *Economic Systems Research*, 13, 109-27.
- Andreff M., W. Andreff & S. Poupaux (2008), Les déterminants économiques de la performance olympique: Prédiction des médailles qui seront gagnées aux Jeux de Pékin, *Revue d'Economie Politique*, 118 (2), 135-69.
- Andreff W. (2001), The Correlation between Economic Underdevelopment and Sport, *European Sport Management Quarterly*, 1 (4), 251-79.
- Andreff W. (2004), Would a Second Transition Stage Prolong the Initial Period of Post-socialist Economic Transformation into Market Capitalism?, *European Journal of Comparative Economics*, 1 (1), 7-31.
- Andreff W. (2006), Sport in developing countries, in W. Andreff & S. Szymanski, eds., *The Handbook on the Economics of Sport*, Cheltenham: Edward Elgar, 308-15.
- Andreff W. (2007), *Economie de la transition: La transformation des économies planifiées en économies de marché*, Paris: Bréal.
- Andreff W. (2009), Comparaison entre les prévisions et les médailles gagnées aux Jeux de Pékin, in INSEP, *Pékin 2008 : Regards croisés sur la performance sportive olympique et paralympique*, INSEP, Secrétariat d’Etat aux Sports, Paris, 241-47.
- Andreff W. & G. Raballand (2010), Is European Football Future to Become a Boring Game?, in W. Andreff, ed., *Contemporary Issues in Sports Economics: Participation and Professional Team Sports*, Cheltenham: Edward Elgar, forthcoming.
- Arellano M. & S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-97.

- Ball D. (1972), Olympic Games Competition: Structural Correlates of National Success, *International Journal of Comparative Sociology*, 13, 186-200.
- Bernard A.B. (2008), Going for the Gold: Who Will Win the 2008 Olympic Games in Beijing, <http://mba.tuck.dartmouth.edu/pages/faculty/andrew.bernard/Beijing2008>.
- Bernard A.B. & M.R. Busse (2004), Who Wins the Olympic Games: Economic Resources and Medal Totals, *Review of Economics and Statistics*, 86 (1), 413-17.
- Boeri T. & B. Severgnini (2009), The Italian Job: Match Rigging, Career Concern and Media Concentration in Serie A, Bocconi University, mimeo.
- Bohlmann H.R. & J.H. van Heerden (2005), The Impact of Hosting a Major Sport Event on the South African Economy, *Working papers series*, n°2005-09, University of Pretoria, Department of Economics.
- Choudhury D.R., P. Bhargava, Reena & S. Kain (2007), Use of Artificial Neural Networks for Predicting the Outcome of Cricket Tournaments, *International Journal of Sports Science and Engineering*, 1 (2), 87-96.
- Clarke S.R. (2000), Home Advantage in the Olympic Games, in G. Cohen & T. Langtry, eds., *Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport, Conference proceedings*, Sydney: University of Technology Sydney, 43-51.
- Coupé T. (2007), Incentives and Bonuses – The Case of the 2006 World Cup, *Kyklos*, 60 (3), 349-358.
- Grimes A.R., W.J. Kelly & P.H. Rubin (1974), A Socioeconomic Model of National Olympic Performance, *Social Science Quarterly*, 55, 777-82.
- Groot L. (2008), *Economics, Uncertainty and European Football. Trends in Competitive Balance*, Cheltenham: Edward Elgar.
- Hawksworth J. (2008), Economic Briefing Paper: Modelling Olympic Performance, PricewaterhouseCoopers, June.
- Heyndels B. & C. Du Bois (2006), Revealed Comparative Advantage and Specialisation in Athletics, European Association for Sport Management conference, Nicosia-Cyprus, *Sport World United*, 50-51, September.
- Hill D. (2009), How Gambling Corruptors Fix Football Matches, *European Sport Management Quarterly*, 9 (4), 411-32.
- Hoffmann R., L.Chew Ging & B. Ramasamy (2002a), Public Policy and Olympic Success, *Applied Economic Letters*, 9, 545-48.
- Hoffmann R., L.Chew Ging & B. Ramasamy (2002b), The Socio-Economic Determinants of International Soccer Performance, *Journal of Applied Economics*, 5, 253-72.
- Horne J. (2004), The Global Game of Football: The 2002 World Cup and Regional Development in Japan, *Third World Quarterly*, 25, 1233-44.
- Houston R.G. Jr & D.P. Wilson (2002), Income, Leisure and Proficiency: An Economic Study of Football Performance, *Applied Economic Letters*, 9, 939-43.
- Jiang M. & L.C. Xu (2005), Medals in Transition: Explaining Medal Performance and Inequality of Chinese Provinces, *Journal of Comparative Economics*, 33 (1), 158-72.
- Johnson D. & A. Ali (2004), A Tale of Two Seasons: Participation and Medal Counts at the Summer and Winter Olympic Games, *Social Science Quarterly*, 85 (4), 974-93.
- Johnson D. & A. Ali (2008). Predictions for Medal Counts at Beijing Olympics, Based on Economic Model, <http://faculty1.coloradocollege.edu/~djohnson/Olympics/Beijing2008predictions.pdf>

- Jolk E., M.J. Karvonen, J. Kihlberg, J. Koskela & J. Noro (1956), Sports in the Cultural Pattern of the World. A Study of the 1952 Olympic Games at Helsinki, Institute of Occupational Health, *mimeo*.
- Kim H.J., D. Gursoy & S.B. Lee (2006), The Impact of the 2002 World Cup on South Korea: Comparison of Pre- and Post-Games, *Journal of Tourism Management*, 27, 86-96.
- Kuper G.H. & E. Sterken, (2008), Evaluation of Beijing 2008 Olympic medal tally forecasts: Who has won? <http://www.eco.rug.nl/~sterken/download/evaluation2008.htm>
- Kurscheidt M. (2006), The World Cup, in W. Andreff & S. Szymanski, eds., *The Handbook on the Economics of Sport*, Cheltenham: Edward Elgar, 197-213.
- Leeds M. & E. Marikova Leeds (2009), International Soccer Success and National Institutions, *Journal of Sports Economics*, 10 (4), 369-90.
- Levine N. (1974), Why Do Countries Win Olympic Medals? Some Structural Correlates of Olympic Games Success: 1972, *Sociology and Social Research*, 58, 353-60.
- Macmillan P. & I. Smith (2007), Explaining International Soccer Rankings, *Journal of Sports Economics*, 8 (2), 202-13.
- Maennig W. & S. du Plessis (2007), World Cup 2010: South African Economic Perspectives and Policy Challenges Informed by Experience of Germany 2006, *Contemporary Economic Policy*, 25 (4), 578-90.
- Maennig W. & Wellebrock C., (2008), Sozioökonomische Schätzungen olympischer Medaillen-gewinne. Analyse-, Prognose- und Benchmarkmöglichkeiten. *Sportwissenschaft 2*, 131-48.
- Monk J. & J. Husch (2009), The Impact of Seeding, Home Continent, and Hosting on FIFA World Cup Results, *Journal of Sports Economics*, 10 (4), 391-408.
- Nevill A., G. Atkinson, M. Hughes & S. Cooper (2002), Statistical Methods for Analyzing Discrete and Categorical Data Recorded in Performance Analysis, *Journal of Sports Sciences*, 20 (10), 829-44.
- Novikov A.D. & A.M. Maximenko (1972), The Influence of Selected Socio-economic Factors on the Levels of Sports Achievements in the Various Countries, *International Review of Sport Sociology*, 7, 27-44.
- Paul S. & R. Mitra (2008), How Predictable Are the FIFA Worldcup Football Outcomes? An Empirical Analysis, *Applied Economic Letters*, 15, 1171-76.
- Pfau W.D. (2006), Predicting the Medal Wins by Country at the 2006 Winter Olympic Games: An Econometric Approach, National Graduate Institute for Policy Studies, Tokyo, January, *mimeo*.
- Poupaux S. & W. Andreff (2007), The Institutional Dimension of the Sports Economy in Transition Countries, in M.M. Parent & T. Slack, eds., *International Perspectives on the Management of Sport*, Amsterdam: Elsevier, 99-124.
- Preuss H., M. Kurscheidt & N. Schutte (2008), Consumption Patterns of Tourists to Sports Mega-Events: Evidence from FIFA World Cup 2006, presented at the 10th IASE Conference, Gijon, May 9-10.
- Rahmann B., W. Weber, Y. Groening, M. Kurscheidt, H.-G. Napp & M. Pauli (1998), *Sozio-ökonomische Analyseder Fussball-Weltmeisterschaft 2006 in Deutschland: Gesellschaftliche Wirkungen, Kosten-Nutzen-Analyse und Finanzierungsmodelle einer Sportgrossveranstaltung*, Cologne: Sport und Buch Strauss.
- Rathke A. & U. Woitek (2008), Economics and the Summer Olympics: An Efficiency Analysis, *Journal of Sports Economics*, 9 (5), 520-37.
- Shibli S. & J. Bingham (2008), A Forecast of the Performance of China in the Beijing Olympic Games 2008 and the Underlying Performance Management Issues, *Managing Leisure*, 13 (3), 272-92.

- Szymanski S. (2002), The Economic Impact of the World Cup, *World Economics*, 3 (1), 169-78.
- Torgler B. (2004), The Economics of the FIFA Football Worldcup, *Kyklos*, 57 (2), 287-300.
- Wang J. & Y. Jiang (2008), Logarithm Model Prediction for 2008 Beijing Olympic Games, *International Journal of Sports Science and Engineering*, 2 (1), 47-51.
- Yamamura E. (2009), Technology Transfer and Convergence of Performance: An Economic Study of FIFA Football Ranking, *Applied Economics Letters*, 16, 261-266.

Appendix 1: Medal wins at the 2008 Beijing Olympics: predictions versus actual outcomes

Countries	Medal wins in Beijing	Hawksworth 2008	Error	Johnson & Ali 2008	Error	Andreff 2008 Lower b.	Andreff 2008 Upper b.	Within*	If not, error	Maennig & Wellenbrock 2008	Error
United States	110	87	-23	103	-7	103	110	yes		95	-15
China	100	88	-12	89	-11	73	86	no	-14	68	-32
Russia	72	79	7	95	16	93	100	no	21	78	6
Great Britain	47	28	-19	28	-19	32	35	no	-12	38	-9
Australia	46	41	-5	26	-20	47	54	no	1	41	-5
Germany	41	43	2	66	25	50	54	no	9	43	2
France	40	30	-10	27	-13	35	38	no	-2	31	-9
South Korea	31	27	-4	24	-7	27	32	yes		27	-4
Italy	28	29	1	29	1	34	36	no	6	29	1
Japan	25	34	9	37	12	37	41	no	12	35	10
Spain	18	19	1	13	-5	20	23	no	2	19	1
Netherlands	16	20	4	22	6	22	24	no	6	20	4
Brazil	15	12	-3	10	-5	10	14	no	-1	10	-5
Hungary	10	17	7	31	21	17	21	no	7	16	6
Poland	10	14	4	24	14	12	16	no	2		
Canada	18	13	-5	17	-1	13	18	yes		14	-4
Turkey	8	11	3	11	3	7	11	yes		10	2
Bulgaria	5	13	8	23	18	10	13	no	5	11	6
Greece	4	15	11	10	6	14	17	no	10	14	10
Mexico	3	8	5	6	3	5	9	no	2	5	2
South Africa	1	8	7	10	9	7	12	no	6		
<i>21 nations</i>	648	636	150**	701	222**	668	764		118**	637***	133
<i>Error in %</i>			23.1%		34.3%				18.2%		20.9%
Ukraine	27	23	-4			24	29	yes		22	-5
Cuba	24	24	0			25	33	no	1		
Belarus	19	15	-4			14	20	yes		15	-4
Kazakhstan	13	9	-4			8	14	yes		8	-5
Romania	8	19	11			19	23		11		

Czech Republic	6	9	3		8	12	no	2	10	4	
Indonesia	5	8	3		1	5	yes		5	0	
Thailand	4	10	6		5	8	no	1			
Iran	2	8	6		2	7	yes				
<i>30 nations</i>	756	761	191**		774	915		133**			
<i>Error in %</i>			25.3%					17.6%			
Kenya	14			2	-12	1	4	no	-10	2	-12
Norway	10			13	3	4	7	no	-3		
New Zealand	9			12	3	2	8	no	-1		
Denmark	7			12	5	6	9	yes		8	1
Switzerland	6			12	6	4	7	yes			
Sweden	5			19	14	6	9	no	1		
Finland	4			13	9	0	3	no	-1	4	0
Nigeria	4			5	1	0	2	no	-2	2	-2
Thailand	4			5	1	5	8	no	-1	7	3
Austria	3			11	8	6	9	no	3	8	5
Belgium	2			11	9	3	5	no	1	5	3
Israel	1			9	8	0	1	yes		4	3
Tunisia	1			5	4	0	0	no	-1		
Ireland	3			12	9	0	1	no	-2	3	0
<i>35 nations</i>	718			842	314**	811	837		168**	749****	180
<i>Error in %</i>					43.7%				23.4%		24.%

* Actual outcome within our confidence interval.

** Absolute value.

*** Over 19 nations

**** Over 33 nations

Appendix 2: National teams cumulative participation to the soccer World Cup semi-finals

Teams	Until 1958	1962	1966	1970	1974	1978	1982	1986	1990	1994	1998	2002	2006
Argentina	1	1	1		1	2	2	3	4	4	4	4	4
Austria	2					2	2		2		2		
Belgium	0			0			0	1	1	1	1	1	
Brazil	3	4	4	5	6	7	7	7	7	8	9	10	10
Bulgaria	0	0	0	0	0			0		1	1		
Chile	0	1	1		1		1				1		
Croatia	0										1	1	1
(ex-)Czechoslovakia*	1	2		2			2		2				2
England	0	0	1	1			1	1	2		2	2	2
France	1		1			1	2	3			4	4	5
Germany	3	3	4	5	6	6	7	8	9	9	9	10	11
Hungary	2	2	2			2	2	2					
Italy	2	2	2	3	3	4	5	5	6	7	7	7	8
Netherlands	0				1	2			2	2	3		3
Poland	0				1	1	2	2				2	2
Portugal	0		1					1				1	2
South Korea	0							0	0	0	0	1	1
Spain	1	1	1			1	1	1	1	1	1	1	1
Sweden	3			3	3	3			3	4		4	4
Turkey	0											1	
USA	1								1	1	1	1	1
Uruguay	3	3	3	4	4			4	4			4	
(ex-)USSR **	0	0	1	1			1	1	1				
(ex-)Yugoslavia***	1	2			2		2		2		2		2

* Czech Republic after 1992 ** Russia after 1991 *** Serbia Montenegro since 1998

A blank means that a team has not participated to the final tournament this year.

Since 1962, a non increasing figure means that a team has participated to the final tournament without reaching the semi-finals.