www.pwc.com/cz

Ice Hockey National Team Performance Model

April 2015

pwc

The consulting firm PwC analysed the conditions of every country for success in the Ice Hockey World Championship that starts May 1 in Prague. Statistical analysis that took into account, besides other things, historical results at championships in the last 20 years, the number of stadiums, the number of ice hockey players registered, demographic and economic indicators or the average annual temperature favours the host - Czech Republic. The silver should go to the Russian team and bronze to the Canadians that will beat Sweden. The dark horse of the Championship will be Switzerland which did not fulfil its potential until now.

In this paper we utilized the regression analysis for estimating the so-called „PwC point index". This index represents historical performance of national teams during the past 20 years (1994 – 2014) of world championships. It also takes into account every place in the final ranking, not just the first three (medal positions) as seen in the table 1.

Table 1: Historical Performance Calculation

Rank	Ranking Points	Medal Bonus	Medal Weight	Final Points
1	16	10	4 * silver team points	192
2	15	10	2 * bronze team points	48
3	14	10	0	24
4	13	0	0	13
5	12	0	0	12
(…)				
16	1	0	0	1

The calculation clearly favors teams that were able to win a medal. We use the following algorithm to recalculate weights of the individual medals – gold is worth 4 silver medals and silver is worth 2 bronze. Unlike similar ranking systems we also graded the rest of the ranking ladder – that means the teams from the 4th place onwards are graded as follows:

Points = Number of teams in total – final ranking of team + 1

In this calculation we also took into consideration the fact that the world championships in the years 1994 – 1997 consisted of only 12 teams taking part in the competition and only from the year 1998 the rules are stable at the amount of 16 teams in the championship pool.

Another relevant factor is the recency of the achieved result – that means how many years ago the corresponding championship took place. A simple reasoning tells us that the older the result is, the less relevant for this analysis it must be. Bearing in mind that we possess 20 year span of data and we need to distribute the points evenly, we chose to edit the calculated points according to appropriate weights. The oldest year was fixated with the weight of 0.33, so the winner of the 1994 season will be awarded only a third of the points the last year's winner would be.

Between the years 2014 (with 100 percent weight) and 1994 (with 33 percent weight) the individual steps were adjusted by 0.0335. Amount of points was multiplied by the corresponding weight and all of the edited point indices for every national team were summed up.

This resulted in the base of the future analysis – dependent variable – the historical performance of the national team

Table 2: Historical Performance of Countries Overview (alphabetical order)

Country	Historical Performance
Austria	32.555
Belarus	71.87
Canada	748.1645
Czech Republic	846.3045
Denmark	50.5765
Finland	532.0255
France	43.456
Germany	81.562
Latvia	85.624
Norway	66.962
Russia	865.8
Slovakia	286.801
Slovenia	10.645
Sweden	711.343
Switzerland	147.3445
United States	161.3235

The next part of this analysis was to identify independent variables that were significant enough. The first obvious source was the data from the *World Economic Outlook* published by the International Monetary Fund. Among the used attributes of above specified countries we chose GDP, unemployment rate, number of inhabitants and others. As the analysis showed later on, none of them was relevant enough to describe the PwC Point Index variable.

As the charts 1 & 2 show, the ability of GDP and number of inhabitants indicators (i.e. economic and demographic criteria) to credibly describe the dependent variable is very low (coefficient of determination is almost zero using the linear function).

This finding implicates that economic prosperity of a country does not influence the performance of the national teams (given the data we used it seems that it is almost contradictory). We decided to abandon these indicators and chose a different set of data where we awaited significantly better performance – statistical data of the International Ice Hockey Federation (IIHF).

Chart 2: Correlation of Historical Performance Score and Number of Inhabitants

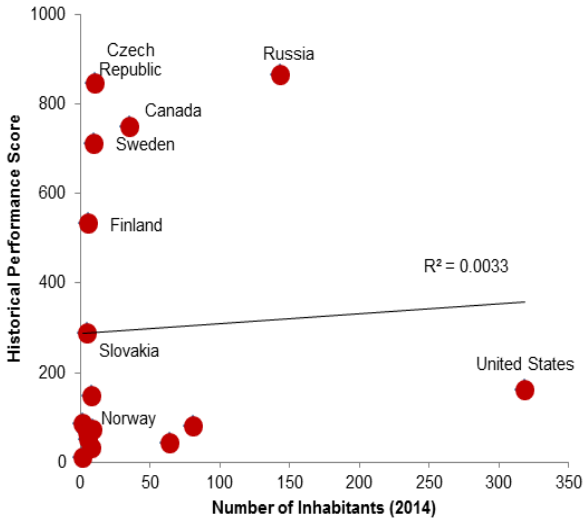
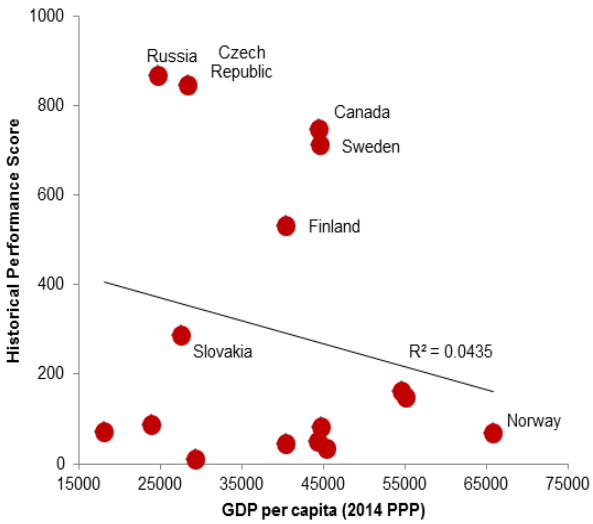


Chart 1: Correlation of Historical Performance Score and GDP per capita



International Ice Hockey Federation presents at its website (http://www.iihf.com/iihf-home/countries/) the member countries data relevant to ice hockey – date of induction to the federation, name of the current president of local ice hockey organization, IIHF ladder country ranking etc. For creation of the model we chose numerical attributes we assumed might describe the PwC Point Index well:

- Number of registered male players
- Number of registered female players
- Number of registered junior players
- Number of registered referees
- Number of indoor rinks
- Number of outdoor rinks

These attributes possess significantly higher ability to describe the performance levels of individual national teams, as shown in the charts 3 and 4.

These variables were then complemented by an attribute describing average air temperature of the countries (data come from the publicly accessible sites on *Wikipedia.org*). We operate under the assumption that countries with colder climate have longer tradition of winter sports such as ice hockey and the sport should be more popular there than in the southern countries. This claim is also supported by the map of current participants of the 2015 world championship.

Figure 1: IIHF WC 2015 Participants

As the last input variable we used an attribute indicating how many times the corresponding country served as a host for the championship in the past 20 years. This attribute further supports the hypothesis that well-performing national teams are mostly in countries with long hockey tradition and popularity.

Chart 3: Correlation of Historical Performance Score and Number of Registered Referees

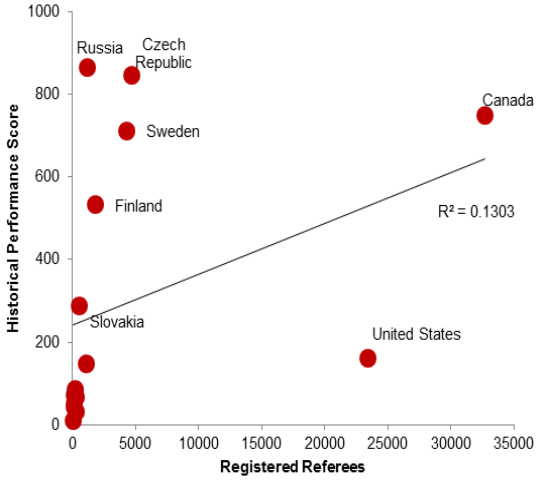
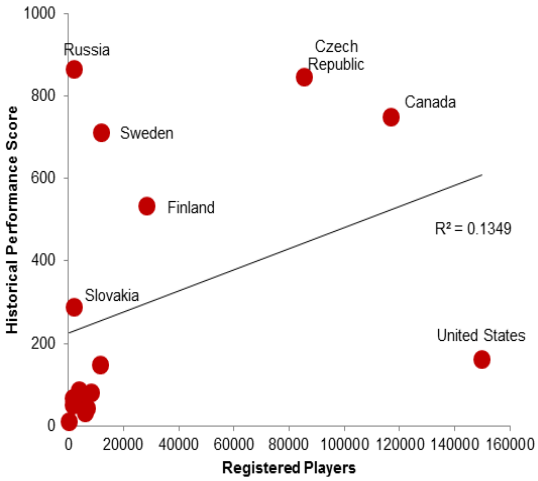


Chart 4: Correlation of Historical Performance Score and Number of Registered Players



The resulting source data are presented in the table 3.

Table 3: Source Data for Regression Analysis (alphabetical order)

Country	Historical Performance	Average Temp.	Count as Host	Registered Players			Referees	Indoor Rinks	Outdoor Rinks
				Male	Junior	Female			
Austria	32.555	7	2	6069	4978	673	265	45	72
Belarus	71.87	5.9	1	3423	4890	42	129	31	3
Canada	748.165	3.6	1	116884	518008	86612	32710	2631	5000
Czech Republic	846.305	6.8	1	85576	22302	2647	4727	158	23
Denmark	50.577	7.5	0	1678	2174	400	111	25	0
Finland	532.026	2.6	3	28589	39263	5830	1821	260	28
France	43.456	11.2	0	6819	10071	1321	109	135	9
Germany	81.562	7.8	2	8389	18084	2898	223	202	45
Latvia	85.624	6	1	3965	1689	86	187	17	0
Norway	66.962	4.4	1	1818	4179	632	270	45	1
Russia	865.8	-0.6	2	1966	81592	712	1164	419	2450
Slovakia	286.801	6.2	1	2122	8675	511	546	64	17
Slovenia	10.645	7.9	0	145	799	76	38	7	0
Sweden	711.343	4.7	3	11884	41521	3434	4296	355	136
Switzerland	147.345	6	2	11584	13740	1091	1106	158	30
US	161.324	11.6	0	149884	302303	67230	23413	1898	500

Regression analysis was performed on this data set using a linear function. All of the input variables were significant enough to be used in the model.

Table 4: Regression Outcome of the Historical Performance Model

Independent Variable	Coefficient	Standard Error	t-Stat
Average Temperature	-6.0149949	6.5930833	-0.91232
Number of Times as a Host	11.889306	27.556902	0.431446
Male Players	0.0066446	0.0013375	4.96801
Junior Players	0.0168595	0.0065651	2.56803
Female Players	-0.1058729	0.019005	-5.57078
Referees	0.0291248	0.0281171	1.035841
Indoor Rinks	0.3514684	0.4023114	0.873623
Outdoor Rinks	-0.2968958	0.1587554	-1.87015

By applying the result back on the original data set (with blank values of the dependent variable) we got a model estimation of team performance according to our input variables. These values represent the performance the teams should accomplish.

Table 5: Application of Regression Analysis on Original Dataset with Dependent Variable Removed (ordered by model estimates)

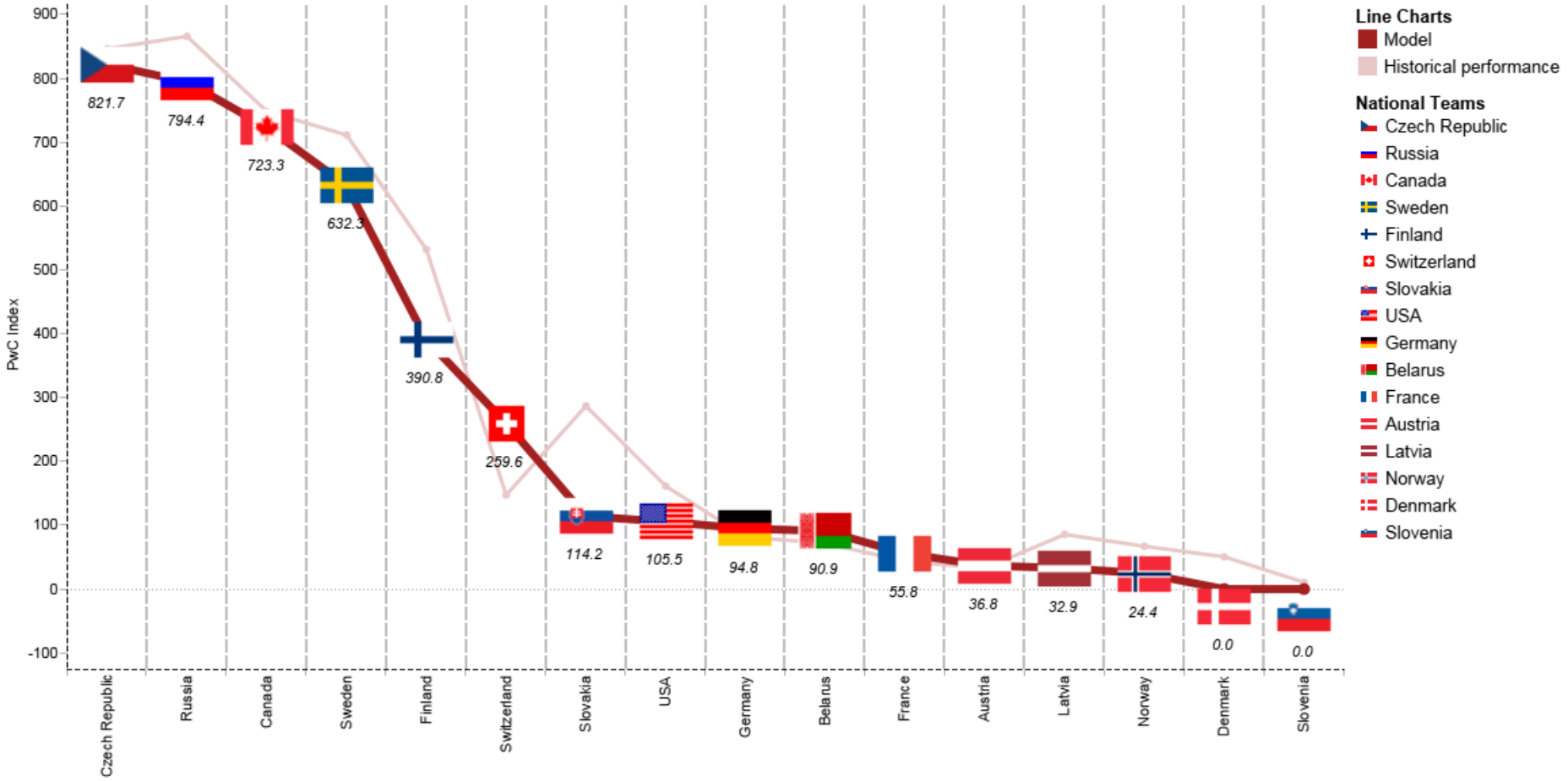
Country	Historical Performance	Model Estimates
Czech Republic	846.3045	821.7345
Russia	865.8	794.4395
Canada	748.1645	723.2646
Sweden	711.343	632.33
Finland	532.0255	390.8101
Switzerland	147.3445	259.6385
Slovakia	286.801	114.1999
United States	161.3235	105.5135
Germany	81.562	94.80114
Belarus	71.87	90.90333
France	43.456	55.82584
Austria	32.555	36.83123
Latvia	85.624	32.93697
Norway	66.962	24.43011
Denmark	50.5765	0
Slovenia	10.645	0

Values for Denmark and Slovenia were adjusted due to the negative prediction of the model, which is absent of any sense for this analysis – even if a country ends up last in the championship, according to our model it is awarded with 1 point (minimal amount). We manually edited Denmark and Slovenia to the value of 0.

Another interesting analysis would be to use similar attributes on a timeline – it would be possible that way to predict the outcome of a championship more precisely. However, the IIHF data source does not contain historical data that we could use for this kind of analysis, so we did not focus on this goal in our work.

The last page contains final visualization of the national ice hockey team performance estimation by regression model described in this paper.

Figure 2: Ice Hockey Teams Performance Model



Contact

Pavel Jankech
Senior manager
PwC Czech Republic
pavel.jankech@cz.pwc.com
+420 251 151 336