Evaluating NHL Point Production Forecasting Methods for Russian Skaters

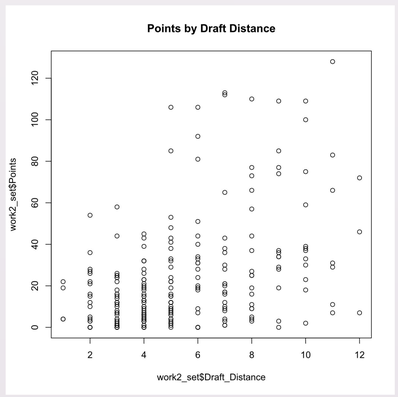
Erica James
Bryn Mawr College

Abstract

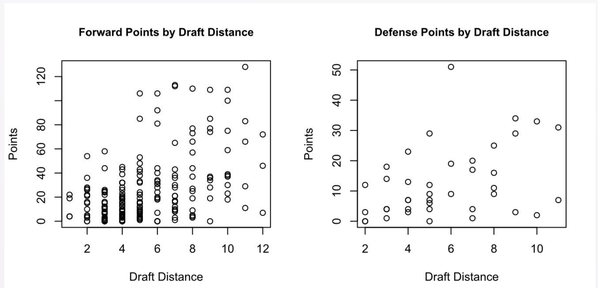
One of the best ways to understand the skills of a prospect is through international tournaments among their age group and skill level as it allows for direct comparison against prospects in more familiar development systems. For recently drafted Russian prospects, this has not been possible for several years for a variety of reasons. This project attempts to simulate the first five years in the NHL of Russian skaters drafted in the first three rounds of the NHL Draft from 2020-2023 by comparing their weighted pre-NHL performance to that of Russian NHL players from 2004-2012 with similar development paths. All data was collected from QuantHockey and analysed in R using the forecast and tidyverse. This resulted in six forecasting models on point production based on position and a logistic regression for likelihood of reaching the NHL. I found that the logistic regression had a McFadden value approaching 1 while the most effective forecasting models for forwards and defenders were within 2 standard deviations from the dataset. This project shows that statistical models can give insight into outcomes from particular development systems in cases where comparative viewing is limited. It does not however, negate the necessity of watching prospects as numbers alone are unable to identify all of the strengths and weaknesses of a particular prospect.

Data and Weighting

- The dataset is made of 57 players drafted in the period 2004-201_ as the training set and is meant to evaluate players drafted from 2020-2023 who have completed their pre-draft development within the Russian system
- The data dataset covers 3 years pre-draft and up to 7 years post draft
- The average player in this set reaches the NHL 2.44 years after drafting
- The dataset overlaps heavily with seasons affected by COVID-19 and thus the point totals are artificially depressed

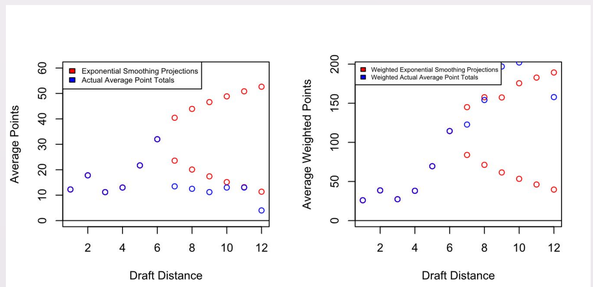


Data and Weighting



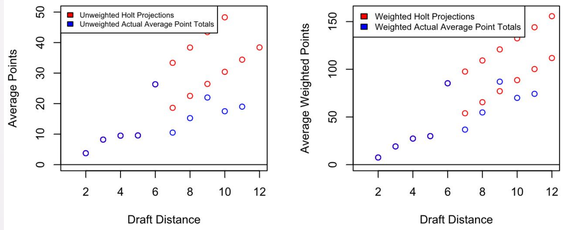
Forecasting Methods

- Five forecasting methods were tested, naïve, TBATS, Holt, ARIMA, and exponential smoothing at both the standard confidence interval and at a confidence level of 75%
- Forecasting analysis was completed on both the weighted and unweighted point totals
- The worst forecasting method for forwards with unweighted points was ARIMA, which produced a constant value of 18.00
- The best method for forwards with unweighted points was exponential smoothing at an 80% confidence interval with a MAPE value of 29.39
- When using the weighted points, exponential smoothing had an MAPE value of 31.10



Forecasting Methods

- The best method for defenders with unweighted points was a Holt at an 80% confidence interval with an MAPE of 61.54
- When using the weighted points, the Holt at an 80% confidence interval had an MAPE value of 73.27
- The average MAPE value for forwards using unweighted points was 33.665 with a standard deviation of 2.854
- The average MAPE value for forwards using weighted points was 33.054 with a standard deviation of 2.940
- The average MAPE value for defenders using unweighted points was 40.367 with a standard deviation of 18.635
- The average MAPE value for defenders using weighted points was 49.092 with a standard deviation of 21.340



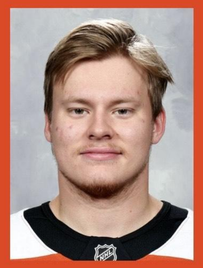
Logistic Regression

- Two logistic regressions were tested, one on likelihood of reaching North American professional hockey and a second on likelihood of reaching the NHL
- The logistic regression with unweighted points on playing North American professional hockey has a McFadden value of 1 and is not statistically significant
- The League_ID variable increases McFadden values but decreases accuracy
- The logistic regression with unweighted points of playing in the NHL has a McFadden value of 0.5 and is statistically significant to 0.01
- The logistic regression of weighted points on playing in the NHL has a McFadden value of 0.5 and is statistically significant to 0.01

Further Steps

- Creating a Shiny app for use by non-technical stakeholders
- Adding a metric to measure "success" through team fit using penalty minutes to assess whether a prospect is playing out of depth
- This model focuses solely on point production and thus disadvantages more defensive players
- Expanding the dataset to the later rounds and the 2024 and 2025 drafts

Matvei Michkov



Profile:

Age: 20

Height:

Weight:

Position: Left Wing

Draft Position: 7th, 2023 NHL Draft

NHL Games Played: 80

NHL Points Per Game: 0.79

NHL PIMs: 46