Jake Cardonick

Optimizing NHL Round 1 Forward Selections: An Improved Draft Model

Objective

This paper aims to uncover two key insights through the application of weighted least squares modeling and machine learning: 1) how NHL teams have historically valued specific statistics—such as league affiliation, goals, assists, physical measurements, and combine metrics—in their first-round draft picks over the past five seasons, and 2) how we can develop a model that not only reflects but outperforms these existing valuation frameworks.

Project Expansion

Initially, the project focused on ordering the value NHL teams place on each junior hockey league. However, recognizing the potential for broader application, I expanded the scope to create a more comprehensive draft model. This model currently includes only first-round forward draft picks and should be primarily applied to players projected to go in the first or early second rounds. Nevertheless, the model has the potential to be expanded to encompass all draft rounds.

Step 1: Current NHL Draft Valuation Model

Parameters

Draft Pick Value: Derived from the "NHL Mock Draft Database." While functional, this aspect of the model could be significantly refined with a more accurate valuation method.

League: The initial focus of the project—identifying the strength of the league in which the player competed during their draft year.

Height, Weight, GPG (Goals Per Game), APG (Assists Per Game), Age

Combine Statistics: I analyzed whether players ranked in the top 10 in various combine metrics (11 metrics in total) using inverse weighted probabilities to normalize the results. Improved access to complete combine data would enhance this aspect of the model.

Combine Attendance: Since no combine was held for 2021 and 2020 prospects, a normalization metric was necessary. Due to limited data, I assumed all first-rounders in other years attended the combine, which presents an area for further improvement.

The Original Draft Model

```
                                                   WLS Regression Results
===================================================================================================================================
Dep. Variable:     Pick Value (https://www.nhlmockdraftdatabase.com/trade-value-chart-2023)   R-squared:                       0.768
Model:                                                                                 WLS   Adj. R-squared:                  0.664
Method:                                                                      Least Squares   F-statistic:                     7.374
Date:                                                                     Tue, 20 Aug 2024   Prob (F-statistic):           1.44e-13
Time:                                                                             12:18:07   Log-Likelihood:                   -inf
No. Observations:                                                                      114   AIC:                               inf
Df Residuals:                                                                           78   BIC:                               inf
Df Model:                                                                               35
Covariance Type:                                                                 nonrobust
============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                   -6363.5102   3282.792     -1.938      0.056   -1.29e+04     172.026
Height (cm)                31.6433     13.590      2.328      0.022       4.587      58.699
Weight (kg)                14.0363      9.330      1.504      0.137      -4.539      32.612
GPG                      1034.0536    248.680      4.158      0.000     538.969    1529.138
APG                      1537.4929    233.891      6.574      0.000    1071.851    2003.134
Age                      -182.1652    152.835     -1.192      0.237    -486.437     122.107
V02 Max                   669.9350    162.507      4.123      0.000     346.409     993.461
Bench                     172.0780    191.218      0.900      0.371    -208.607     552.763
Pull-Ups                  205.8825    133.503      1.542      0.127     -59.901     471.666
Grip L_R                  -42.3326    135.493     -0.312      0.756    -312.079     227.413
Agility L_R                23.1095    134.544      0.172      0.864    -244.747     290.966
Duration                 -723.9943    200.595     -3.609      0.001   -1123.348    -324.640
Mean Power                155.0500    182.474      0.850      0.398    -208.228     518.328
Peak Power                344.9389    219.722      1.570      0.120     -92.494     782.372
Fatigue                  -183.6587    188.558     -0.974      0.333    -559.048     191.731
Vert J                    571.1597    164.032      3.482      0.001     244.597     897.722
Horz J                   -507.7283    213.001     -2.384      0.020    -931.780     -83.677
Had_Combine              -210.7492    151.608     -1.390      0.168    -512.578      91.080
League_CZECHIA            3.13e-12   5.34e-13      5.867      0.000    2.07e-12    4.19e-12
League_DEL               3342.8631    525.981      6.355      0.000    2295.715    4390.011
League_Denmark            807.0290    530.311      1.522      0.132    -248.740    1862.798
League_HockeyAllsvenskan 3644.4871    602.330      6.051      0.000    2445.341    4843.633
League_J20               1899.2719    426.559      4.453      0.000    1050.059    2748.485
League_KHL               3443.7265    571.871      6.022      0.000    2305.219    4582.234
League_LIIGA             3067.5288    477.157      6.429      0.000    2117.582    4017.476
League_MHL               2168.8989    490.441      4.422      0.000    1192.507    3145.291
League_NCAA              2936.8350    463.015      6.343      0.000    2015.043    3858.627
League_NTDP              1952.4352    400.412      4.876      0.000    1155.276    2749.595
League_OHL               1972.3857    382.444      5.157      0.000    1210.999    2733.772
League_PHL               1429.8704    611.296      2.339      0.022     212.875    2646.866
League_QMJHL             2080.7730    419.918      4.955      0.000    1244.781    2916.765
League_SHL               3973.8480    512.599      7.752      0.000    2953.341    4994.355
League_SL                3507.4410    545.741      6.427      0.000    2420.955    4593.927
League_Slovakia          3166.9442    572.120      5.535      0.000    2027.941    4305.947
League_USHL              1963.0218    436.372      4.499      0.000    1094.272    2831.771
League_VHL               2979.9325    622.418      4.788      0.000    1740.794    4219.071
League_WHL               2083.4444    380.198      5.480      0.000    1326.529    2840.360
```

The Improved Draft Model

```
                                                   WLS Regression Results
===================================================================================================================================
Dep. Variable:     Pick Value (https://www.nhlmockdraftdatabase.com/trade-value-chart-2023)   R-squared:                       0.886
Model:                                                                                 WLS   Adj. R-squared:                  0.838
Method:                                                                      Least Squares   F-statistic:                     18.63
Date:                                                                     Tue, 20 Aug 2024   Prob (F-statistic):           2.70e-21
Time:                                                                             12:21:00   Log-Likelihood:                   -inf
No. Observations:                                                                       93   AIC:                               inf
Df Residuals:                                                                           65   BIC:                               inf
Df Model:                                                                               27
Covariance Type:                                                                 nonrobust
============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                   -6310.4188   3212.336     -1.964      0.054   -1.27e+04     105.060
Height (cm)                34.8233     11.821      2.946      0.004      11.215      58.431
Weight (kg)                37.1348     10.235      3.628      0.001      16.695      57.575
GPG                       799.4643    231.627      3.452      0.001     336.874    1262.055
APG                      1643.7518    205.973      7.980      0.000    1232.394    2055.109
Age                      -288.7630    166.684     -1.732      0.088    -621.654      44.128
V02 Max                   484.0870    122.619      3.948      0.000     239.200     728.974
Duration                 -854.4430    173.602     -4.922      0.000   -1201.151    -507.735
Peak Power                258.9282    168.380      1.538      0.129     -77.350     595.206
Vert J                    282.2361    154.287      1.829      0.072     -25.896     590.368
Horz J                    180.0585    196.770      0.915      0.364    -212.918     573.035
Had_Combine                43.3120    119.406      0.363      0.718    -195.159     281.783
League_CZECHIA          -7.576e-12   9.62e-13     -7.879      0.000    -9.5e-12   -5.66e-12
League_DEL               2786.8318    350.641      7.948      0.000    2086.554    3487.109
League_HockeyAllsvenskan 2833.7808    473.988      5.979      0.000    1887.162    3780.399
League_J20               1208.9470    256.596      4.711      0.000     696.489    1721.405
League_KHL               3034.1517    411.518      7.373      0.000    2212.293    3856.011
League_LIIGA             2451.4372    342.471      7.158      0.000    1767.474    3135.400
League_MHL               1721.0594    426.606      4.034      0.000     869.069    2573.050
League_NCAA              2667.4275    285.553      9.341      0.000    2097.139    3237.716
League_NTDP              1116.3312    263.757      4.232      0.000     589.573    1643.089
League_OHL               1370.4473    258.476      5.302      0.000     854.235    1886.659
League_PHL                 36.6035    358.640      0.102      0.919    -679.650     752.857
League_QMJHL             1580.9561    283.885      5.569      0.000    1013.999    2147.914
League_SHL               3460.5932    318.211     10.875      0.000    2825.082    4096.104
League_SL                3098.2616    461.490      6.714      0.000    2176.603    4019.920
League_USHL              1302.7062    362.167      3.597      0.001     579.409    2026.004
League_VHL               2322.3844    434.088      5.350      0.000    1455.451    3189.318
League_WHL               1523.0558    242.184      6.289      0.000    1039.381    2006.731
```

Players excluded from the updated model: Holtz, Boucher, Gaucher, Wood, Holloway, Coronato, Mesar, Honzek, Amirov, Rosen, Schaefer, Barlow, J_Perreault, Othmann, Stramel, Brisson, Lucius, Wiesblatt, Bourgault, Olausson, and Stillman. **These players were excluded because their production decreased in the years following their draft year.**

A few combine metrics were also excluded due to high p-values. Yet, the R-squared and adjusted R-squared values both still improved.

Purpose of the Models

The original draft model serves as a reflection of how NHL teams have made their first-round draft selections over the past five years, incorporating all available data on player statistics, physical attributes, combine metrics, and the leagues in which players competed. This model is intended to capture the raw decision-making processes of teams, without filtering for post-draft performance. It represents the baseline of how teams have historically valued various factors at the time of drafting.

In contrast, the improved draft model is designed to refine this analysis by excluding players whose production significantly declined after being drafted. These players are identified as busts—individuals whose post-draft performance did not meet expectations. The purpose of this model is to explore how draft outcomes might improve if teams had been able to identify and avoid selecting these players. By comparing the improved model to the original, we can identify which factors are more strongly associated with long-term success in the NHL and which factors may have led teams to overvalue certain players during the draft process.

Comparing the Models

The R-squared value increased from 0.768 in the original model to 0.886 in the improved model, indicating that the latter explains 88.6% of the variance in draft pick value, compared to 76.8% in the original model. Similarly, the adjusted R-squared value rose from 0.664 to 0.838, further confirming the improved model's superior explanatory power even after adjusting for the number of predictors. The F-statistic in the improved model more than doubled (from 7.374 to 18.63), demonstrating a much stronger overall significance, and the probability of the F-statistic (p-value) decreased substantially, reinforcing the strength of the relationships between the predictors and the outcome.

Goals vs. Assists

In the original draft model, the ratio of GPG to APG coefficients is approximately 0.673, indicating that goals per game are weighted somewhat less heavily than assists per game. This suggests that NHL teams might have placed more importance on a player's ability to contribute to goals through assists rather than purely through goal-scoring in their draft evaluations.

However, in the improved draft model, this ratio decreases to approximately 0.487, further emphasizing the higher value placed on assists relative to goals when underperforming players are excluded. This larger disparity in the improved model suggests that successful draft outcomes are more closely associated with players who are strong playmakers (as indicated by a higher APG) rather than those who primarily score goals (as indicated by GPG).

Age

In the original draft model, the coefficient for age is -182.1652 with a p-value of 0.237, suggesting that while younger players might be valued slightly more highly, the effect is not statistically significant. This indicates that, at the time of drafting, NHL teams may have considered age as a factor, but it did not play a crucial role in their decision-making process.

However, in the improved draft model, the age coefficient becomes more negative, increasing in magnitude to -288.7630, and the p-value drops to 0.088. Although the p-value is still slightly above the conventional threshold for statistical significance, the trend is clear: younger players are associated with higher draft pick values, particularly when excluding those players who did not live up to expectations post-draft.

Height and Weight

When comparing the significance of height and weight between the original and improved draft models, it becomes evident that these physical attributes gain considerable importance in predicting draft pick value once underperforming players are excluded. In the original model, the coefficient for height is 31.6433 with a p-value of 0.132, indicating that height was not statistically significant in determining draft outcomes. Similarly, the weight coefficient was 14.0636 with a p-value of 0.137, also lacking statistical significance. However, in the improved model, both height and weight become more prominent predictors. The height coefficient increases to 34.8233, and its p-value drops dramatically to 0.004, while the weight coefficient rises to 37.1348, with its p-value also falling to 0.001. This substantial decrease in p-values highlights that, once busts are excluded, taller and heavier players are more strongly associated with higher draft pick values, indicating that these physical attributes are more reliable indicators of long-term success in the NHL than originally perceived by teams.

Combine Metrics

The analysis of the combine metrics—V02 Max, Duration, Peak Power, Vert J (Vertical Jump), and Horz J (Horizontal Jump)—reveals important insights into their significance and impact on draft outcomes between the original and improved draft models. V02 Max, a key measure of cardiovascular endurance, is significant in both models, underscoring its importance as a predictor of NHL success. V02 Test Duration (if they were top 10) shows a more pronounced negative coefficient in the improved model (-854.4430) compared to the original (-723.0943), highlighting that players are often overdrafted when successful in this metric. Peak Power is positive and while not quite statistically significant, it shows promise. Vert J, a measure of explosive leg power, decreases in both coefficient and significance in the improved model, suggesting that vertical jump height might have been overvalued by teams. Finally, Horz J shows the most dramatic shift, with its coefficient changing from significantly negative in the original model to a positive but insignificant value in the improved model. This indicates that players are not being overdrafted based on horizontal jump. Rather, it simply isn't significant in either direction. Together, these findings suggest that while physical fitness and athleticism are crucial, teams may need to refine their focus on which specific metrics best predict long-term success in the NHL.

Leagues

Based on the improved draft model, the leagues can be ranked in terms of perceived difficulty and competitiveness as follows: At the top are the SHL (Swedish Hockey League) and Liiga (Finnish Elite League), which are considered the most competitive and likely to produce NHL-ready talent. Following closely are the HockeyAllsvenskan and the DEL (German League), which also rank highly, indicating strong competition and solid player development. The KHL (Kontinental Hockey League) and NCAA (National Collegiate Athletic Association) are next, reflecting their ability to produce valuable players. The MHL (Russian Junior Hockey League) and QMJHL (Quebec Major Junior Hockey League) occupy the middle tier, offering good but not top-tier competition. Further down the list are leagues like the SL (Swiss League), OHL (Ontario Hockey League), and WHL (Western Hockey League), which are still competitive but less so than those ranked higher. The USHL (United States Hockey League), NTDP (National Team Development Program), J20 (Swedish Junior Hockey League), and VHL (Russian Supreme Hockey League) follow, with the PHL (Polish Hockey League) and CZECHIA (Czech Extraliga) rounding out the list as the least competitive in this ranking.

Testing the 2015 Draft Class

To test the 2015 draft class, I outputted the model's value for each (first round) forward in the class. I then ranked the model outputs 1-21. After, I ranked the players by PPG from the 2023 season to help compare.

Player	Actual Rank	Model Rank	PPG Rank
McDavid	1	1	1
Eichel	2	2	4
Strome	3	3	10
Marner	4	7	3
Zacha	5	13	11
Meier	6	4	12
Rantanen	7	6	2
Crouse	8	9	14
Gurianov	9	10	17
Debrusk	10	21	15
Senyshyn	11	19	21
Barzal	12	14	5
Connor	13	16	6
E. Svechnikov	14	8	18
Eriksson Ek	15	4	9
White	16	17	20
Boeser	17	11	7
Konecny	18	20	8
Roslovic	19	18	13
Beauvillier	20	15	16
Merkley	21	12	19

Players with a 5 or greater difference between model rank and actual rank (where they were drafted among forwards):

1. **Zacha**: Model win. Solid player, but a bad draft pick that high, nonetheless.
2. **Debrusk**: Right in the middle. Overdrafted but not the worst player on this list as the model predicted.
3. **Senyshyn**: Model win.
4. **E. Svechnikov**: Draft win. Big miss from the model here. Obviously, his brother ended up being a star (not that it matters).
5. **Eriksson Ek**: Model win.
6. **Boeser:** Model win.
7. **Beauvillier**: Model win.
8. **Merkley**: Draft win.

Sources for Improvement

To enhance the accuracy and robustness of the draft model, several key areas for improvement can be identified. First, expanding the sample size by incorporating more draft classes would provide a broader dataset, allowing for more generalizable insights and reducing potential biases that might arise from a smaller sample. Additionally, refining the model used to determine draft pick value is crucial; the current model relies on data from the "NHL Mock Draft Database," but developing a more precise and nuanced valuation framework could lead to better predictions. Another area for improvement is the definition of what constitutes a "bust." A more sophisticated method for identifying and excluding busts from the analysis could improve the model's ability to differentiate between players who underperform and those who meet or exceed expectations. Moreover, having access to more comprehensive combine data would allow for a deeper analysis of how specific physical and athletic metrics correlate with long-term NHL success. Lastly, refining the normalization process for combine participation, especially for years where the combine did not occur, would help ensure that all players are evaluated on a level playing field. Addressing these areas would significantly enhance the model's predictive power and its utility for NHL teams during the draft process.