The Hockey Project: An analysis of Forwards

Saakshi Mehta

29882157

ECON 490: Seminar in Applied Economics

University of British Columbia

Abstract

This paper examines the likelihood of a forward to get drafted into National Hockey League. This study focuses on two important relationships. Firstly, what performance statistics determine the probability of an eligible CHL forward to get drafted. Secondly, how does this probability vary across players in different years of eligibility. Using Probit regression analysis, the results estimate that a unit change in goals per game and assists per game increases the draft probability by 33% and 13% respectively. Additionally, data suggests that year of eligibility matters significantly and youngest players have the highest probability of getting drafted.

Keywords: Ice Hockey, Entry Drafts, Linear Probability Model, Probit Model

The Hockey Project: An analysis of Forwards

Hockey enthusiasts across North America eagerly wait for the annual National Hockey League (NHL) entry draft at the end of each season. On the basis of draft results, they form expectations about how their team would perform in the next season. Across North American sports leagues, annual entry draft or "the draft" is a common means of selecting junior players and signing them to national teams. Eligible junior players are selected primarily on the basis of their performance in junior league along with evaluations conducted by Central Scouting Service (CSS). In Ice hockey, on-ice performance of a players is commonly measured through goals, assists, points and penalty minutes. Performance statistics also include per game averages like goals per game, assists per game or points per game. In order to secure a position in entry draft, junior hockey players are expected to have superior performance statistics and outstanding CSS evaluations. But this is not necessarily true for all positions.

Existing literature on hockey statistics agree that on-ice performance statistics are not a perfect measure to predict players' future performance or value as they do not fully capture the true performance of the variety of roles in the team (Dawson and Magee, 2001; Tingling, Masri and Martell, 2011). For example, defensemen do not contribute to the game by scoring goals but rather they aim to engage in more defensive strategies and contribute more to the team performance. Although new metrics have been developed and introduced over time to overcome these challenges, the reliability of such measures have not been established (Tingling, Masri & Martell, 2011). Even though these statistics are not a perfect measure, but these statistics are reliable indicators of future performance of the young athletes.

The key assumption of this study is that performance statistics are positively related to a player's likelihood of getting drafted. The rationale for this assumption is that higher performance statistics is reflective of a player's superior skills on ice and sincere dedication towards the game. Under this assumption, forwards are more likely to get selected by NHL teams as they tend to have higher performance statistics compared to their counterparts. Compared to defensemen, forwards are more likely to accumulate higher points through goals and assists due to their offensive strategy and proximity to the net. Defensemen, on the other hand, are more likely to have lower points than forwards but they are better known for their engagement in defensive strategies and physical durability. Their performance could probably be better estimated through Penalty Minutes, Plus Minus and other physical or performance measures. Research also supports the notion that forwards provide greater value to the team performance, relative to defensemen. Researchers Chan, Cho and Novati (2012) found that goaltenders tend to provide the most value to team performance (on per-game basis), followed by forwards and then defensemen. In Table 1, I have collected NHL entry draft data from 2011 to 2019 that provides evidence for the theory discussed above. Since 2011, forwards constitute at least 53% of overall drafts. This indicates that NHL teams have a preference for forwards over Defensemen. In 2019, out of 217 hockey players drafted in 2019 NHL entry drafts, 129 were Forwards. This constitutes 60% of the players who got drafted overall. According to CHL stats, 71 junior players were drafted, and 46 of them were Forwards (65%).

Table 1: Percentage of Forwards drafted in National Hockey League from 2011-19

Year	Total Forwards Drafted in NHL	% of Forwards Drafted Overall	Forwards Drafted from CHL	% of Forwards Drafted from CHL
2019	129 (217)	59%	46 (71)	65%
2018	118 (217)	54%	39 (78)	50%
2017	117 (217)	54%	55 (89)	62%
2016	118 (211)	56%	55 (96)	58%
2015	112 (211)	53%	54 (95)	57%
2014	124 (210)	59%	62 (95)	65%
2013	122 (211)	58%	60 (101)	59%
2012	110 (211)	52%	59 (99)	60%
2011	112 (211)	53%	64 (101)	63%

Note: This data has been collected and compiled from www.chl.ca. It provides number of Forwards drafted to NHL from 2011-2019. This table also specifies number of Forwards drafted from CHL during the same time period. The percentage of forwards has been calculated as the number of forwards (NHL and CHL) divided by total number of overall drafts (NHL and CHL) respectively. The figure in brackets in the total number of players drafted in that year.

Researchers Tingling, Masri and Martell (2011) note that Games Played (GP) is a good measure of long-term performance and it is easy to use and comparable across positions. In their study, Tarter et al (2009) used *Sports Performance Index for Hockey (SPI-H)* as a composite index that incorporates measures of physical fitness, hockey game statistics at the junior level along with the Scout evaluations to determine the likelihood of the long-term career in NHL. This provides further evidence that junior league performance statistics along with other measures are a reliable indicator of long-term performance and increased likelihood of getting into and staying in NHL. Therefore, the focus of this paper is to utilize these game statistics and its transformations to determine the likelihood of getting drafted for Forwards.

Given that Forwards are more likely to get drafted in NHL and their game statistics are reliable indicators of their performance, I investigate what performance statistics can be used to determine the likelihood of a Forward getting drafted to NHL. CKM Sports Management provided player-by-player data on performance statistics and personal attributes of all the players eligible for 2019 Entry Draft from Canadian Hockey League (CHL). The data constitutes the performance statistics of junior players who have played for Western Hockey League (WHL), Quebec Major Junior Hockey League (QMJHL) and Ontario Hockey League (OHL). The data includes quantitative performance statistics – games played, goals, assists, penalty minutes, plus/minus along with qualitative data – eligibility year (calculated from birth year), position on ice, junior league, and team changes. As the dependent variable in binary – either a player gets drafted or not, I employ Probit model and Linear probability model to determine key performance statistics that would influence a player's likelihood of getting drafted.

The results of this study are novel as the probability of a unit increase in goals per game and assist per game is quantified. Using marginal effects of Probit regression, I find that the goal per game increases the probability of getting drafted by 33% while assists per increases the probability by approximately 13%. Another important contribution of this paper is that the focus is solely on Forwards. To my knowledge, there is no current research work that uses performance statistics to inform the likelihood of Forwards getting drafted.

There is also a statistically significant effect of eligibility year on the likelihood of getting drafted. Junior players in their first year of eligibility are more likely to get drafted than players in the second or third year of eligibility. The interaction of eligibility year with performance statistics show the same pattern. In other words, an increase in performance statistics has a larger effect on the predicted draft probability for younger players relative to older players. Baker and Logan

(2007) found similar results that concluded that relatively younger players were drafted earlier. These findings were later confirmed by Fumarco et al who explored the long-term impact of relative age effect (RAE) among the NHL players on player productivity. RAEs occur when those who are relatively older for there are more likely to succeed. These have been more frequently with Canadian Ice Hockey. Fumarco et al found that even though the players from the third and fourth quarter born are under-represented, but they are more likely to end up with higher salaries and other performance statistics. This finding can be valuable to predict the performance of players in NHL.

In addition to age of a player, other characteristics like player's mobility across teams also have an impact on performance and earnings of a player. Researchers Vincent and Eastman (2012) found that cumulative effect of moving across teams on earnings is negative. However, the effect of mobility with respect to performance statistics has not been addressed. In this paper, I would attempt to identify the relationship between a player's mobility and the likelihood of getting drafted to NHL. This will be covered in the Discussion section along with other robustness checks. As part of the robustness checks, another binomial model– *Logit regression* has been examined. The results of logit regression confirm the results with slight variation in coefficients. These results do not affect the main findings of the study at large. These analyses can be used by trainers, mentors and instructors to guide the players to work on specific skills required to increase the number of average goals or assists per game.

An article by MIT Sloan Sports Analytics (2019) throws light on some of the challenges that sports analysts face while working with Hockey data and how that poses as a bottleneck in team's decision-making process. Hockey players are draft eligible at the age of 18, which is younger than most sports leagues. This poses as a challenge as "it's harder to predict results for young athletes", the article notes. Researchers Tarter et al echo the same sentiment and reassert that selecting junior hockey players is a gamble. They noted that the probability of accurately identifying an athlete's long-term career in NHL drops to the probability of "predicting a coin flip."

The primary empirical challenge of the study is the high correlation between the variables. Higher numbers of games played are correlated with more ice time and hence, more points (goals and assists). This can cause multicollinearity between variables and bias the coefficients. Another challenge to this model is that the data is cross sectional. This could cause the coefficients to suffer from omitted variable bias as the model cannot control for individual differences between players. Moreover, exclusion of a player's physical attributes, scouting evaluations and cognitive strength can further bias the coefficient. Lastly, this research solely focuses on the performance statistics. Tarter et al (2009) demonstrate that performance statistics explain only a part of what shapes the overall performance of a player. As noted earlier, future research can use a composite index or a combination of physical attributes and performance statistics can be used to resolve these issues. The models used in the current study are adequate to answer basic questions of strategy: how many points per game or number of games played in junior league will yield the highest probability of getting drafted.

The remainder of the paper is organized as follows. Section II discusses the setting of Ice Hockey and Annual Entry Drafts. Section III presents the data. Section IV describes the empirical model. Section V presents the results. Section VI discusses the implications of the results and shortcomings of the model. Conclusions are provided in Section VII.

II. Ice Hockey and Annual Entry Drafts

Ice Hockey is a multifaceted game played between two teams, where a team consists of nine players at a time – three forwards, two defensemen and one goaltender. Both teams try to maximize the number of goals by shooting the puck past the goaltender and into the net. The team that scores most goals is the winner. This game is played in three, 20-minute rounds. Goals are awarded to players who scores the goal. An assist in awarded to player who facilitates the goal either through shooting, passing or deflecting the puck. It can be awarded to a maximum of two players, excluded the player who scored the goal. Points for each player are calculated as the total of goals and assists (Riley, 2017).

Unlike other major sports leagues, hockey players tend be drafted relatively earlier. Junior hockey players become eligible for the draft between the ages 18-20. For 2019 NHL draft, athletes born between January 1, 1999, and September 15, 2001 were eligible. This means that these players only have three chances to get drafted in their “dream league.” The motivation of this paper lies in the three chances that these players get to make it to NHL. These entry drafts are turning points for young hockey players. Most of these young hockey players have dedicated themselves to train and master their skills for this “make-or-break” moment.

Annual entry drafts are common across North American sports leagues where teams have the opportunity to recruit elite junior players from the pool of eligible players. This recruitment mechanism intends to increase competitiveness by strengthening the weaker teams and diminishing the strength of high performing teams (Booth, 2004). These entry drafts usually take place in the off season where the teams get to pick junior eligible players. Annual draft is broken down into seven rounds, where 31 NHL teams pick elite junior hockey players in a sequential order. Typically, the worst performing team enter a lottery and get to select the best players from the eligible pool. This allows worst-performing teams to select before the better-performing teams and add players that would improve their future performance.

It is intriguing to know whether there is a systematic way to determine the probability of getting drafted in NHL based on the players' performance in the junior leagues and their eligibility year. As these players get drafted at such an early age, it can be challenging to predict their future performance based on their limited data from Junior League. The goal of this research is to help coaches and mentors support young hockey players in realizing their dream. To support this process, it is integral to know what attributes of the players can determine their chances of realizing their dreams. By identifying these indicators, coaches and mentors can support the athletes and help them to focus on developing a more comprehensive skillset.

III. Data

A. Data Description and Transformations

Data for this research has been provided internally by CKM Sports Management team, community partner of University of British Columbia. Data provided by CKM Sports Management provides player-by-player data for CHL players who were eligible for NHL Entry Draft 2019. The data constitutes information of athletes who have played for Western Hockey League (WHL), Quebec Major Junior Hockey League (QMJHL) or Ontario Hockey League (OHL).

CHL statistics were individual leagues were compiled together and cross referenced with NHL data to find the eligible junior hockey players. Data constituted the following information - player name, birth year, junior league, position, current team, team changes, draft order and team drafted to. Their performance statistics included Games Played, Goals, Assists, Points (sum of Goals and Assists), Penalty Minutes and Plus/Minus. The raw metrics were transformed to per game averages. To see the effects of per game statistics, I transform the variables Goals and Assists by dividing them by games played during the season. *Goals per game* are calculated as the fraction of total games played (Total goals scored/ Total number of games played). Similarly, *Assists per game* are calculated as the fraction of total games played (Total assists scored/ Total number of games played). Table 2 shows the descriptive statistics of the key game metrics.

Table 2: Summary Statistics of key performance measures

Variable	Mean	Std.Dev.	Min	Max
Games Played	53.436	16.804	1	70
Goals	10.208	10.482	0	53
Goals per Game	.175	.17	0	1.24
Assists	17.045	14.337	0	78
Assists per Game	.293	.227	0	1.421
Points	27.252	23.372	0	117
Points per Game	.467	.367	0	1.92
Penalty Minutes	34.003	24.828	0	149
Plus/Minus	-.16	19.173	-77	76

Note: This table shows summary statistics of key performance statistics of the current sample.

Independent Variables

Performance statistics analyzed for forwards include the following raw metrics and per game statistics. Table 3 summarizes and describes all the specifications used in this study.

Table 3: Description of independent variables

Variable	Description
Games Played	Total number of games played by a Forward
Goals	Total number of times a forward hits the puck into the net
Goals per Game	Calculated as number of goals divided by number of names played (Goals/Games Played)
Assists	Total number of times a forward assists or enables a goal by either passing, hitting or deflecting the puck
Assists per Game	Calculated as number of assists divided by number of names played (Assists/Games Played)
Points	Sum of Goals and Assists
Points per Game	Calculated as points divided by number of names played (Points/Games Played)

Note: This table describes the predictor variables used in this study.

Dependent Variable

Drafted is the binary dependent variable in this study. This variable assumes the value of 1 if the CHL forward has been drafted in the 2019 entry draft. Otherwise, it takes the value 0. While using the linear regression model and non-linear probit model, this variable is interpreted as the probability of a CHL forward getting drafted in NHL Entry Draft.

B. Data setup

The data for individual CHL leagues was compiled together to form master data set for all CHL junior players. These observations were cross-referenced and merged with the NHL data that included data for CHL players who got drafted in 2019 draft. After merging the data on player's name and team, the data had 1030 observations *(n = 1030).* A dummy variable *Forward* was generated that took the value of 1 if the role of the player was either center, left-wing or right-wing.

Forward = 1, if player was a forward and 0, otherwise.

Another dummy variable *TeamChange* was also generated that took the value of 1 when the player had played for more than one team.

TeamChange = 1, if player switched teams and 0, otherwise.

This data included defensemen, goaltenders and European players who played less than 10 games in CHL. All the data was included in the preliminary analysis. As the focus of this study is on determining the predicted probability of forwards, defensemen and goalies were eventually removed from the data. This data also included Europeans players who played for CHL briefly. Since they played few games in season, their inclusion distorted the data. Hence, players who played less than 10 games were removed.

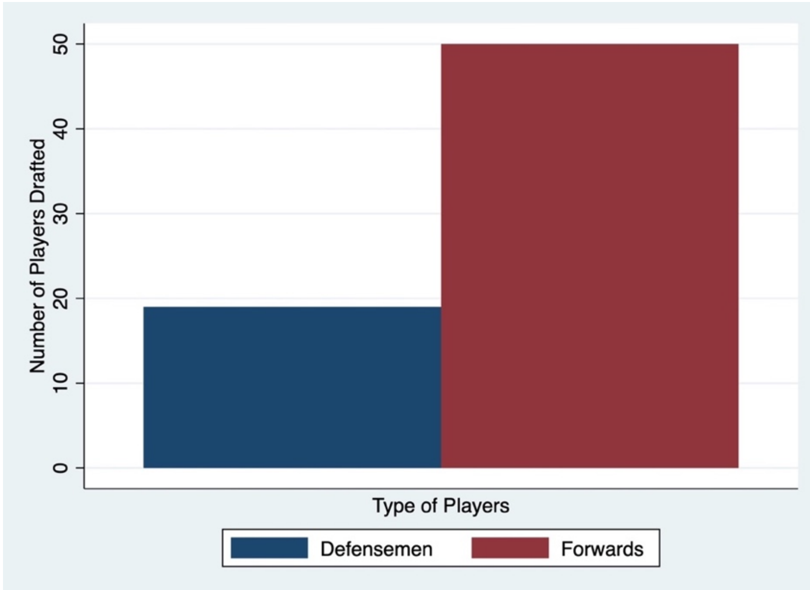


Figure 1. This figure displays the number of Defensemen drafted from Canadian Hockey League in National Hockey League Entry Draft 2019

C. Preliminary Analysis

Once the data was prepared for analysis, there were 668 draft eligible forwards from CHL (n = 668). Out of 668 forwards, 46 were drafted in 2019. Figure 1 illustrates that 46 forwards were drafted compared to 19 Defensemen. Table 4 shows the differences in averages of key performance statistics between drafted forwards and forwards who did not get drafted. Drafted forwards have 16% higher Games Played, 117% higher Goals, 100 % higher assists and over 107% higher points than Non-Drafted Forwards.

Table 4 – Summary Statistics for Drafted Forwards and Not Drafted Forwards

Variable	Forwards Drafted *Mean*	Forwards Not Drafted *Mean*	Percent Difference
Games Played	62.435 (7.884)	53.614 (17.07)	16%
Goals	27.217 (10.621)	12.537 (10.662)	117%
Goals per Game	.431 (.15)	.215 (.173)	100%
Assists	34.435 (11.636)	17.217 (14.568)	100%
Assists per Game	.551 (.172)	.297 (.232)	86%
Points	61.652 (19.627)	29.754 (24.207)	107%
Points per Game	.983 (.272)	.513 (.379)	92%
Penalty Minutes	38.435 (24.347)	31.465 (23.62)	22%

Note: This table shows the summary statistics for Forwards who got drafted in 2019 and compares it with the average performance statistics of forwards who did not get drafted. Percentage difference for the average differences are calculated in the last column.

IV. Empirical Model

The following equation estimates the predicted probability of getting drafted based on raw performance statistics:

$$Pr\ (Drafted = 1) = \beta_0 + \beta_1\ GamesPlayed_i + \beta_2\ Goals_i + \beta_3\ Assists_i + \beta_4\ EligibilityYear_i + \beta_5\ (Forward = 1) + \varepsilon_{i,} \quad (1)$$

where Pr*(Drafted =1)* is the probability of a junior player to get drafted into the 2019 NHL, $GamesPlayed_i$ is the number of games played in the CHL junior league, $Goals_i$ are the number of times the puck is shot past the goalie and into the net, $Assists_i$ are the number of times a player facilitated the goal without actually scoring the goal. This can be awarded to a maximum of two players, not including the player who scored the goal. A goal can be facilitated through shooting, deflecting or passing the puck. Recall that points are the sum of goals and assists (Points = Goals + Assists). Therefore, points are excluded from this specification since goals and assists are already included as regressors. This is done to avoid potential bias from multicollinearity. The alternative equation has been addressed in the robustness checks. $EligibilityYear_i$ is calculated on the basis of a player's birth year. For example, a player born in 2001 (18 years old) is in their first year of eligibility. Based on the birth year, the Eligibility Year = 1. *(Forward = 1)* is a dummy variable that takes the value of 1one when the players is a forward. As noted earlier, the dummy variable takes the value 1, when the player plays either one of the following positions: Center, Left Wing or Right Wing.

The original equation is modified by including per game average in place of raw game statistics. Transformations of goals and assists are used by including - *Goals per game* and *Assists per game*:

$$\Pr(Drafted = 1) = \beta_0 + \beta_1\, GamesPlayed_i + \beta_2\, Goalspergame_i + \beta_3\, Assistspergame_{i.} \quad (2)$$
$$+ \beta_4\, EligibilityYear_i + \beta_5 (Forward = 1) + \varepsilon_{i,}$$

Equations 1 and 2 are tested using a linear model - Linear probability model (LPM) and a non-linear model - Probit regression.

V. Results

A. Linear Probability Model (LPM)

Linear probability model assumes linear relationship between the regressors and outcome variable. Table 5 presents the results of LPM. According to regression results, higher number of games played is negatively correlated with higher likelihood of getting drafted to NHL. This can be interpreted as – for every game played, the probability of getting drafted would decline by 0.2%. Although the coefficient is small but it is significant at 5% significance level. This is counter intuitive because higher number of games played is correlated with higher points. This should increase the draft probability! However, this could be true because of the non-linear relationship between games played and probability of getting drafted. After controlling for non-linearity by adding - *Games Played Squared,* the coefficient on Games Played is still negative but it is not significant anymore. The coefficient on Goals and Assists are positive and statistically significant across both regressions. Column 3 presents the likelihood of getting drafted on the basis of per game averages, while controlling for the non-linear relationship of games played.

Figure 2 and *Figure 3* illustrate the linear prediction of probability for games played by and points respectively. Although, these graphs show that likelihood of getting drafted increases with higher games played and points. But these graphs predict the probability of getting drafted to be *less than 0* for both specifications i.e., Pr(Drafted = 1) < 0. This does not make sense as the probability of getting drafted cannot be less than 0 or more than 1. These illustrations demonstrate that LPM does not explain the data very well. To overcome this challenge, I employ the Probit regression that models the probability of getting drafted using the cumulative standard normal distribution function. In this model, the predicted probability is never below 0 or above 1 (Stock and Watson, 2011).

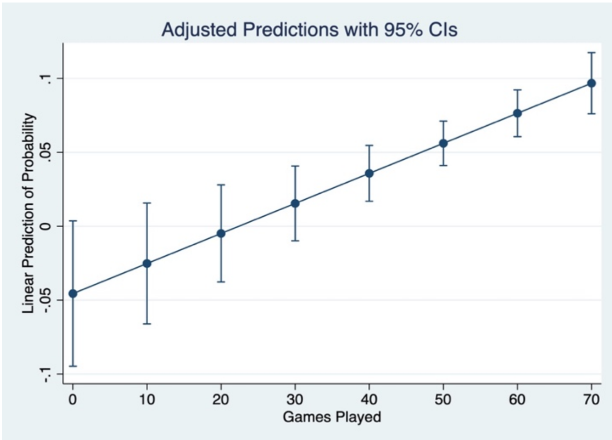


Figure 2

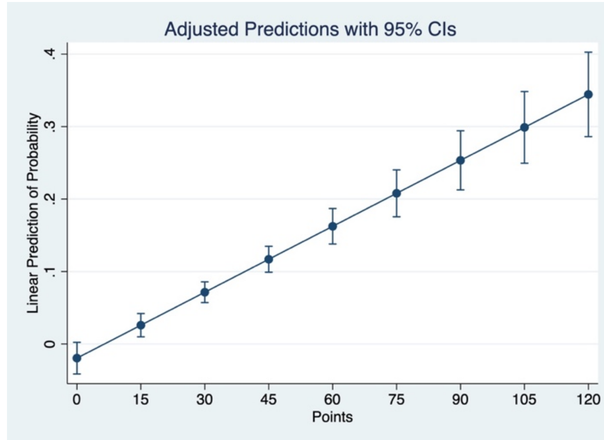


Figure 3

Figure 2-3. These figures present the linear prediction of probability based on games played by an athlete and the points earned by the players respectively. These figures show how the linear probability models predicts the probability less than 0 in both cases.

B. Probit Regression Model

Since Probit Model is non-linear, the slope coefficients cannot be easily interpreted (Stock and Watson, 2011). However, signs of coefficients indicate the relationship between variables. Accordingly, higher goals and assists increase the likelihood of getting drafted in Column 4 and 5. Players in second and third year of eligibility have a disadvantage of getting drafted. Like LPM, higher games played is negatively related with likelihood of getting drafted. But this effect disappears when the non-linear relationship of Games played has been accounted for in Column 5. Column 6 presents the likelihood of getting drafted on the basis of per game averages, while controlling for the non-linear relationship of games played.

Table 5: Predicted Probabilities of Getting Drafted to National Hockey League based on Performance Statistics (Goals and Assists) and Eligibility Year

	(1) LPM_Raw	(2) LPM_Raw	(3) LPM PerGame	(4) Probit_Raw	(5) Probit _Raw	(6) Probit PerGame
Games Played	-0.002**	-0.002	-0.004	-0.017	0.037	0.131*
	(0.001)	(0.002)	(0.002)	(0.011)	(0.069)	(0.079)
Games Played (Squared)		0.000 (0.000)	0.000 (0.000)		-0.001 (0.001)	-0.001 (0.001)
Goals	0.009***	0.009***		0.103***	0.105***	
	(0.001)	(0.001)		(0.019)	(0.019)	
Goals per Game			0.389***			5.404***
			(0.076)			(1.011)
Assists	0.002**	0.002**		0.032***	0.031***	
	(0.001)	(0.001)		(0.011)	(0.011)	
Assists per Game			0.178***			2.174***
			(0.057)			(0.638)
Eligibility Year						
EligYear = 2	-0.182***	-0.182***	-0.183***	-3.070***	-3.071***	-2.952***
	(0.021)	(0.021)	(0.021)	(0.507)	(0.505)	(0.482)
EligYear = 3	-0.232***	-0.232***	-0.228***	-3.481***	-3.494***	-3.266***
	(0.023)	(0.023)	(0.023) (0.057)	(0.491)	(0.490)	(0.451) (0.638)
Constant	0.129***	0.135***	0.105**	-1.983***	-3.065*	-7.094***
	(0.031)	(0.048)	(0.049)	(0.571)	(1.651)	(2.080)
Obs.	668	668	668	668	668	668
Pseudo R^2	.z	.z	.z	0.555	0.558	0.551

Standard errors are in parenthesis

**** p<0.01, ** p<0.05, * p<0.1*

Note: This table presents the regression results for both models along with different specifications.

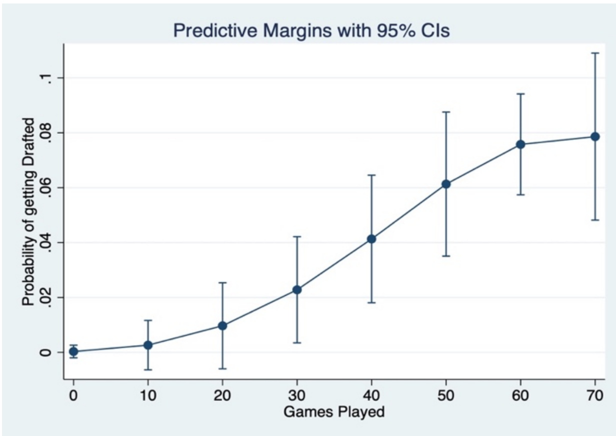


Figure 4

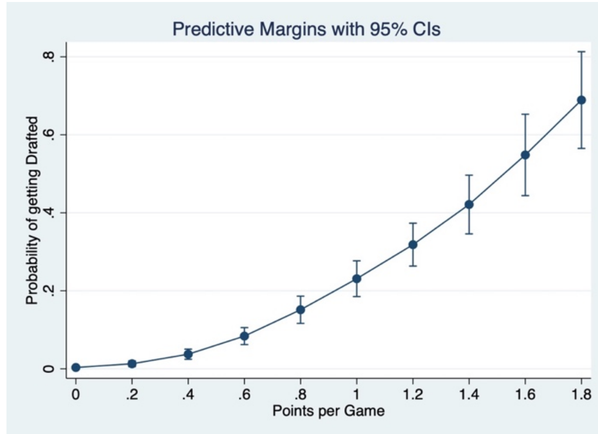


Figure 5

Figure 4-5. These figures plot predicted probability of getting drafted with respect to games played and points per game

Figures 4-7 display probability of getting drafted predicted by Probit model with respect to games played, points per game, goals per game and assists per game respectively. In *Figure 4* the probability increases linearly until a player has played 60 games, but it starts flattening out for games higher than 60. *Figure* illustrates increasing "returns" to probability as draft probability increases with points per game. *Figure 6* exhibits the non-linear behavior of goals per game. The "S curve" of Goals per game demonstrates that the likelihood of getting drafted increases with higher goals per game. The probability peaks at 100% when goals per game is equal to 1.8 (*Goals per game = 1.8*). *Figure 7* shows that the probability of getting drafted peaks at 30% when assist per game is equal to 1.4 (A*oals per game = 1.8*).This suggests that players derive more benefits from scoring higher goals compared to scoring assists. This is not surprising because goals contribute more to the team performance and this sends a positive signal to NHL team managers.

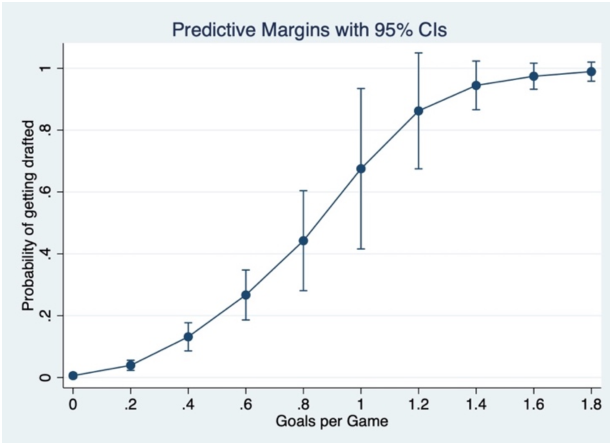


Figure 6

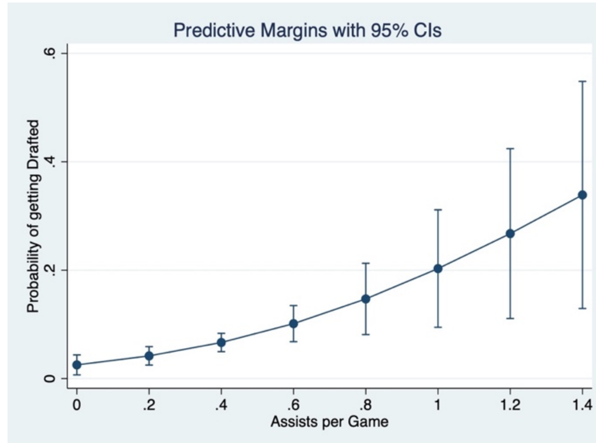


Figure 7

Figure 6-7. These figures plot predicted probability of getting drafted with respect to goals per game and assists per game.

Table 6: Marginal Effects on Predicted Probabilities using Probit model

	Delta-method					
	dy/dx	Std.Err.	z	P>z	[95%Conf.	Interval]
Games Played	0.001	0.001	0.510	0.609	-0.002	0.003
Goals per game	**0.332**	0.056	5.950	0.000	0.223	0.442
Assists per game	**0.134**	0.038	3.540	0.000	0.060	0.208
Eligibility Year						
EligYear = 2	-0.249	0.025	-9.750	0.000	-0.299	-0.199
EligYear = 3	-0.256	0.024	-10.590	0.000	-0.304	-0.209

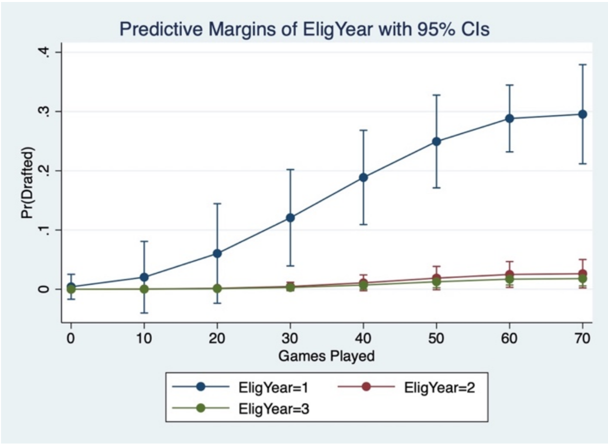
Note: dy/dx for factor levels is the discrete change from the base level.

Using the marginal effects in Table 6, coefficients can be interpreted in a meaningful way. On per-game basis, an increase in goals increases the draft probability of a forward by 33% whereas every assist increases the probability by 13%. These results also quantify the disadvantage of being in second or third year of eligibility. For athletes who are not in first year of eligibility, the likelihood of getting drafted drops by 25%, holding other performance statistics constant. These results are statistically significant at 5% significance level.

VI. Summary and Discussion

In this paper, I have examined two important relationships. Firstly, I analyze what performance statistics are important to predict the probability of an eligible CHL forward to get drafted. Secondly, I look at how this probability varies across players from different years of eligibility. Using marginal effects of Probit regression, I find that every goal per game increases the probability by 33% while every assist per game increases the probability by 13%. This effect is especially large for athletes who are in the first year of their eligibility. This is consistent with the findings of researchers Baker and Logan (2007) who reported that relatively younger players were drafted earlier for Canadian and North American hockey players. One of the explanations for this phenomenon could be that younger players are expected to be more productive (Deaner, Lowen and Cobley, 2013). Given the limited data available on junior players, youngers player with high performance statistics signal more sincerity and dedication towards the game at a relatively younger age. This positive signal might indicate a longer and promising future in NHL, which can enhances the probability of a junior player to be recruited by NHL teams. *Figures 8-11* illustrate the effect of interaction of a player's performance statics and eligibility year on predicted probability of being selected.

A. Interaction Results



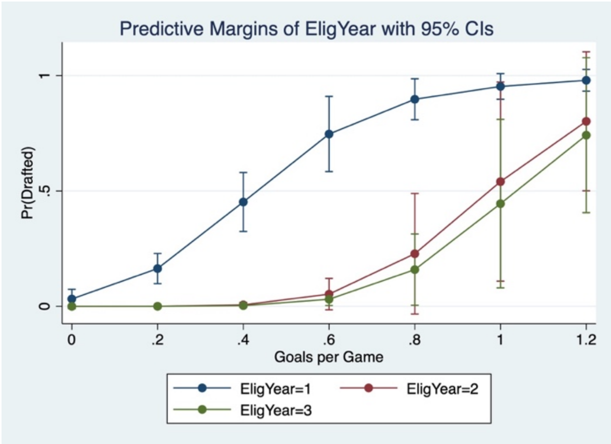


Figure 8 *Figure 9*

Figure 8-9. These figures present predicted probability of getting drafted for different groups of eligibility with respect to games played and goals per game.

Games Played. In *Figure 8*, the predicted probability of getting drafted increases with more games, especially for athletes in first year of eligibility. The increase in probability peaks at 30% for younger player whereas it barely changes the probability for older athletes. This increasing effect visible for games played greater than 20 and less than 60. Players who played more than 60 games see a relatively lower increase in probability. However, this effect is missing for players in their second and third year of eligibility. They do not seem to derive any benefit from playing more games.

Goals per game. Similarly, *Figure 9* displays that younger players have clear advantage as their probability of getting drafted is close to 100% for goals per game greater than 1. Interestingly, this difference between the players across different years of eligibility is greatest between 0.6 to 0.8 but this difference starts diminishing when goals per game is greater than 1. This effect disappears as goals per game exceeds 1. After this point, all players have same probability of getting drafted. An explanation for this is that players with high goals per game indicate a high value to the team for every game they play. This essentially means that players with higher goals per game are performing extremely well in the number of games that they played. This could reflect exceptional skills of the player, which could explain why this effect disappears.

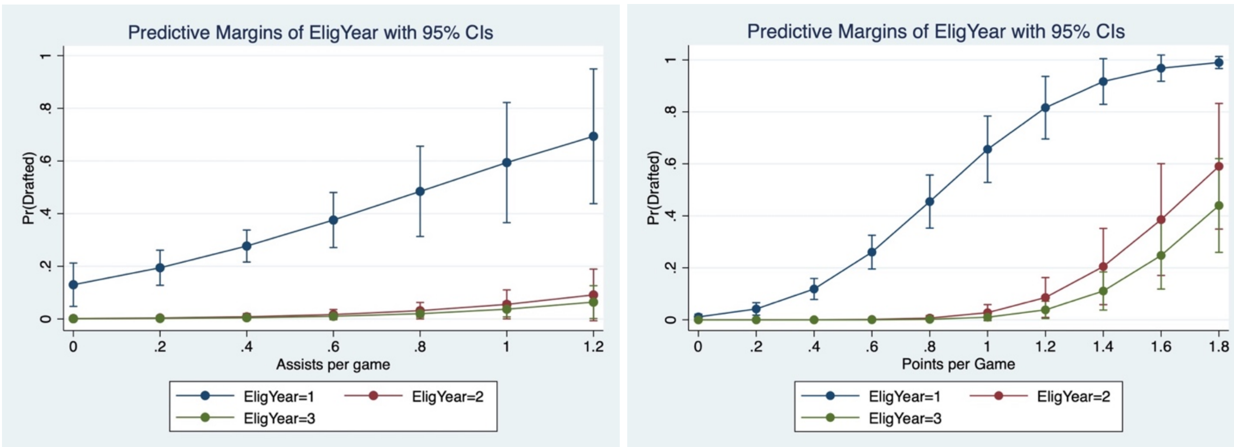


Figure 10 *Figure 11*

Figure 10-11. These figures present predicted probability of getting drafted for different groups of eligibility with respect to assists per game and points per game.

Assists per game. *Figure 10* shows the linear relationship of assists per game and draft probability for younger players. Older players do not derive any advantage of increasing their assists per game as their probability of being drafted lies between 0 to 5 per cent irrespective of increase in assists per game. However, younger players still continue to benefits from every assist per game. For assist per game equal to 1, the likelihood of getting drafted lies between 60 to 70%.

Points per game. *Figure 11* represents the combined advantage for players in first year of eligibility. As points are the sum of goals and assists earned by players, points per game shows a combined effect of goals per game and assists per game. These results show that younger athletes enjoy a higher probability of being drafted and they accrue higher benefits from every point they earn during their junior league.

B. Potential model – Logit Regression

To check for the robustness of the results, Logit regression model has been used to test the robustness of findings. Logit model models the probability of dependent variable = 1, as the cumulative standard logistic distribution function. Table 7 presents the results of Probit and Logit regression (Stock and Watson, 2011). Column 1 and 2 provide regression results for raw game statistics that include games played, goals and assists using the original specifications. Column 3 and 4 run the same regression using per game averages – goals per game and assists per game.

Table 7: Regression results with Probit and Logit

	(1) Probit Raw	(2) Logit Raw	(3) Probit PerGame	(4) Logit PerGame
Games played	0.037	0.074	0.131*	0.247
	(0.069)	(0.134)	(0.079)	(0.156)
Games Played Squared	-0.001	-0.001	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Goals	0.105***	0.195***		
	(0.019)	(0.037)		
Goals per Game			5.404***	9.761***
			(1.011)	(1.911)
Assists	0.031***	0.057***		
	(0.011)	(0.020)		
Assists per Game			2.174***	4.035***
			(0.638)	(1.220)
Eligibility Year				
EligYear = 2	-3.071***	-6.074***	-2.952***	-5.870***
	(0.505)	(1.105)	(0.482)	(1.066)
EligYear = 3	-3.494***	-6.670***	-3.266***	-6.182***
	(0.490)	(0.997)	(0.451)	(0.902)
			(0.638)	(1.220)
Constant	-3.065*	-5.572*	-7.094***	-12.979***
	(1.651)	(3.207)	(2.080)	(4.104)
Obs.	668	668	668	668
Pseudo R^2	0.558	0.557	0.551	0.551

Standard errors are in parenthesis
**** p<0.01, ** p<0.05, * p<0.1*

Again, since this model is non-linear, the slope coefficients cannot be easily interpreted. Therefore, table 7.1 provides the marginal effects of Logit model to provide a more meaningful interpretation. Both models provide the same signs for the included regressors with some differences in the coefficients. Logit model predicts that higher goal per game and assist per game increases the probability of getting drafted for Forwards by 32% (31.5%) and 13% respective. These results are very similar to the original results predicted by Probit model where the goal per game and assist per game increases the likelihood of draft by 33% and 13% respectively.

Table 7.1: Marginal Effects of Goals and Assists per game on predicted probabilities using Logit model

	dy/dx	Delta-method Std.Err.	z	P>z	[95%Conf.	Interval]
Games Played	0.001	0.001	0.420	0.678	-0.002	0.003
Goals per game	**0.315**	0.055	5.720	0.000	0.207	0.423
Assists per game	**0.130**	0.038	3.430	0.001	0.056	0.205
Eligibility Year						
EligYear = 2	-0.260	0.025	-10.220	0.000	-0.310	-0.210
EligYear = 3	-0.264	0.024	-10.980	0.000	-0.311	-0.217

Note: dy/dx for factor levels is the discrete change from the base level.

C. *Alternative Specification – Points and Points per game*

Previously, the equations included Goals and Assists as specifications to replace points to avoid multicollinearity. In this part, points and points per game are included as the specifications. The alternative equations including Points and Points per game are as follows:

$$\Pr(Drafted = 1) = \beta_0 + \beta_1\, GamesPlayed_i + \beta_2\, Points + \beta_3\, EligibilityYear_i + \beta_5\, (Forward = 1) + \varepsilon_{i,} \quad (3)$$

$$\Pr(Drafted{=}1) = \beta_0 + \beta_1\, GamesPlayed_i + \beta_2\, Pointspergame + \beta_3\, EligibilityYear_i + \beta_5\, (Forward = 1) + \varepsilon_i, \quad (4)$$

Table 8 presents the regression results of Probit and Logit model using equations 3 and 4. The signs of all the variables are same as before. This indicates that there alternating the specification does not alter the relationship between the key variables.

Table 8: Predicted Probabilities of Getting Drafted to NHL based on Points and Points per game

	(1) Probit_ Raw	(2) Probit PerGame	(3) Logit_ Raw	(4) Logit PerGame
Games Played	-0.010	0.030**	-0.020	0.051**
	(0.011)	(0.012)	(0.022)	(0.021)
Points	0.056***		0.102***	
	(0.007)		(0.014)	
Points per Game		3.315***		6.114***
		(0.415)		(0.803)
Eligibility Year				
EligYear = 2	-2.870***	-2.850***	-5.702***	-5.743***
	(0.475)	(0.465)	(1.041)	(1.035)
EligYear = 3	-3.114***	-3.042***	-6.070***	-5.925***
	(0.437)	(0.419)	(0.914)	(0.873)
Constant	-2.139***	-4.494***	-3.729***	-7.981***
	(0.592)	(0.809)	(1.185)	(1.478)
Obs.	668	668	668	668
Pseudo R^2	0.530	0.534	0.529	0.536

Standard errors are in parenthesis
*** *p<0.01*, ** *p<0.05*, * *p<0.1*

To quantify the effect of points per game on the draft probability, table 8.1 and 8.2 show the marginal effects of Probit and Logit model respectively. It is important to note that holding everything else constant, predicted probability of getting drafted is approximately 21% now. In other words, a unit change in points per games increases the draft probability by 21%. Recall that for a unit change in goals per game and assists per game, the predicted probability was 33% and

13%. The average effect of those probabilities is captured by marginal effect of points per game at 21%. This result in consistent with the information presented earlier in *Figure 11*.

Table 8.1: Marginal Effects of Points per game on Predicted Probabilities using Probit model

		Delta-method				
	dy/dx	Std.Err.	z	P>z	[95%Conf.	Interval]
Games Played	0.002	0.001	2.530	0.011	0.000	0.003
Points per game	0.213	0.021	10.340	0.000	0.172	0.253
Eligibility Year						
EligYear = 2	-0.247	0.026	-9.550	0.000	-0.298	-0.196
EligYear = 3	-0.252	0.025	-10.130	0.000	-0.301	-0.203

Note: dy/dx for factor levels is the discrete change from the base level.

Table 8.2: Marginal Effects of Points per game on Predicted Probabilities using Logit model

		Delta-method				
	dy/dx	Std.Err.	z	P>z	[95%Conf.	Interval]
Games Played	0.002	0.001	2.460	0.014	0.000	0.003
Points per game	0.205	0.021	9.950	0.000	0.164	0.245
Eligibility Year						
EligYear = 2	-0.260	0.026	-10.140	0.000	-0.311	-0.210
EligYear = 3	-0.262	0.025	-10.670	0.000	-0.311	-0.214

Note: dy/dx for factor levels is the discrete change from the base level.

D. Added Specifications – Team Change

Another way to test the robustness of the findings is by adding the specification of team change. As discussed earlier, existing literature suggest that mobility of the players can influence the probability of getting drafted (Vincent and Eastman, 2012). To check for that, the same regressions were run using the Team Change variable. Team change is the dummy variable that takes the value of 1 when players have played in more than one teams during their junior league.

$$\Pr\ (Drafted = 1) = \beta_0 + \beta_1\, GamesPlayed_i + \beta_2\, Goalspergame_i + \beta_3\, Assistspergame_i$$

$$+ \beta_4\, EligibilityYear_i + \beta_5 TeamChange + \beta_6 (Forward = 1) + \varepsilon_i$$

Table 9 presents the regression results with the additional specification. The results show that the coefficients and significance of the key variables do not vary a lot. This shows that the effect of team change on the probability is very minimal. Table 9.1 confirms this by providing the marginal effects of probit model with added specification.

Table 9: Regression results with Team Change Specification

	(1) LPM	(2) Probit
Games Played	-0.003	0.130
	(0.002)	(0.080)
Games Played Squared	0.000	-0.001
	(0.000)	(0.001)
Goals per Game	0.388***	5.298***
	(0.076)	(1.010)
Assists per Game	0.176***	2.210***
	(0.057)	(0.639)
Eligibility Year		
EligYear = 2	-0.182***	-2.965***
	(0.021)	(0.484)
EligYear = 3	-0.224***	-3.214***
	(0.023)	(0.455)
TeamChange = 1	-0.023	-0.404
	(0.024)	(0.417)
Constant	0.102**	-7.022***
	(0.049)	(2.094)
Obs.	668	668
Pseudo R^2	.z	0.554

Note: This table presents the regression results of Linear Probability model and probit model using team mobility as an added specification.

Table 9.1: Marginal Effects of Team Change on Predicted Probabilities using Probit model

	dy/dx	Delta-method Std.Err.	z	P>z	[95%Conf.	Interval]
Games Played	0.001	0.001	0.520	0.606	-0.002	0.003
Goals per game	0.323	0.056	5.810	0.000	0.214	0.432
Assists per game	0.135	0.037	3.610	0.000	0.062	0.208
Eligibility Year						
EligYear = 2	-0.246	0.025	-9.680	0.000	-0.295	-0.196
EligYear = 3	-0.251	0.025	-10.250	0.000	-0.300	-0.203
TeamChange	-0.022	0.020	-1.100	0.273	-0.062	0.017

Note: dy/dx for factor levels is the discrete change from the base level

E. Limitations

Even though the findings of this research are robust. But this research faces empirical challenges that were discussed throughout the paper. The primary empirical challenge of the study is the high correlation between the variables. Higher numbers of games played are correlated with more ice time and hence, more points (goals and assists). This can cause multicollinearity between variables and bias the coefficients. The problem of multicollinearity can be removed by replacing the current specifications with better game metrics or by using instrumental variables. Another challenge to this model is that the data is cross sectional. This could cause the coefficients to suffer from omitted variable bias as the model cannot control for individual differences between players. The challenges presented by cross sectional data can be overcome by using longitudinal data. This can also resolve any potential omitted variable bias caused by individual differences between the players. Lastly, there are some obvious omitted variables because of exclusion of a player's physical attributes, scouting evaluations and cognitive strength which can further bias the coefficient. Existing research on development of athletes have identified a range of environmental and genetic factors that can directly or indirectly influence the performance of players (Baker and

Logan, 2007). Therefore, any research is incomplete with including physical fitness index, scouting data from CSS, and more comprehensive game statistics.

VII. Conclusion

Following the robustness checks using different model and using alternative specifications, main findings of the remain largely intact. For forwards, the probability of getting drafted to NHL is highest when -

1. *Players are in the first year of eligibility*. Athletes in the first year of eligibility have a significant advantage over athletes in second or third year of eligibility. In fact, players in second or third year of eligibility are 25% less likely to be drafted for same performance statistics.
2. *Goals per game is equal to or greater than 1.2*. Goals are more important than assists in determining the likelihood of getting drafted. An additional goal increases the probability of getting drafted by 33% whereas an additional assist increases the probability by only 13% (per-game basis).
3. *Games played is equal to or greater than 60.* Forwards who play more games tend to have higher points but it is important to note that the increase in games should be complemented by high goals or assists. If the per game averages decrease, this could send a negative signal to NHL team managers and scouting agents and lower the likelihood of getting drafted.

The results of the given can be summarized in the following sentence –

"Holding individual differences constant, if a CHL forward has played more than or equal to 60 games, with goals per game higher than or equal to 1 (or Points per Game greater than or equal to 1.4), then the predicted probability of getting drafted is 1 on average."

From these findings, it is clear that players with higher performance statistics have a higher probability of being drafted, especially in the first year of their eligibility. Therefore, forwards should leverage their position on ice to accumulate higher points in junior league to secure a position in the dream league. To conclude, the quantitative analysis of performance statistics and player attributes can provide insight into what should be the ideal number of points (goals, assists) that a forward should accumulate before NHL entry draft. The models used in the current study are adequate to answer basic questions: how many points per game, games played in the junior league will yield the highest probability of getting drafted. And what will be the impact of each goal or assist on the probability of getting drafted. However, the individual differences between the players and their skills cannot be captured by these analyses. In other words, these results work on average, holding individual differences constant. With the recent development in machine learning and artificial intelligence techniques, more dynamic models can be used to overcome the limitations of research in Hockey.

References

An Analytics Approach to the NHL Entry Draft. (MIT Sloan Sports Analytics Conference). Retrieved from: http://www.sloansportsconference.com/mit_news/analytics-approach-to-the-nhl-entry-draft/?fbclid=IwAR1TCpAPZ86K1gsGvbWiji4VQfUAn6uFmeHyZNGfSzEraXCtuU-EO2lJNw4q

Baker, J., & Logan, A. J. (2007). Developmental contexts and sporting success: Birth date and birthplace effects in national hockey league draftees 2000–2005. *British Journal of Sports Medicine, 41*(8), 515-517. doi:10.1136/bjsm.2006.033977

Booth, R. (2004). the economics of achieving competitive balance in the australian football league, 1897–2004. *Economic Papers: A Journal of Applied Economics and Policy, 23*(4), 325-344. doi:10.1111/j.1759-3441.2004.tb00986.x

CHL Players selected in NHL Draft (2011-2019). Retrieved from: https://chl.ca/article/71-chl-players-selected-in-2019-nhl-draft

Chan, Cho, J. A., & Novati, D. C. (2012). Quantifying the contribution of NHL player types to team performance. Interfaces, 42(2), 131-145. doi:10.1287/inte.1110.0612

Dawson, D., & Magee, L. (2001). The national hockey league entry draft, 1969-1995: An application of a weighted pool-adjacent-violators algorithm. *The American Statistician, 55*(3), 194-199. doi:10.1198/000313001317098158

Deaner, R., Lowen, A., & Cobley, S. (2013). Born at the wrong time: Selection bias in the NHL draft. *Plos One, 8*(2), e57753. doi:10.1371/journal.pone.0057753

Fumarco, L., Gibbs, B. G., Jarvis, J. A., & Rossi, G. (2017). The relative age effect reversal among the national hockey league elite. PLoS One, 12(8)

Riley, S. N. (2017). Investigating the multivariate nature of NHL player performance with structural equation modeling. PloS One, 12(9), e0184346. doi:10.1371/journal.pone.0184346

Stock, J. H., & Watson, M. W. (2011). *Introduction to econometrics* (3rd ed.). Boston: Addison-Wesley.

Tarter, B., Kirisci, L., Tarter, R., Jamnik, V., Gledhill, N., & McGuire, E. J. (2009). Use of the sports performance index for hockey (SPI-H) to predict NHL player value. *International Journal of Performance Analysis in Sport, 9*(2), 238-244. doi:10.1080/24748668.2009.11868480

Tingling, P. M., Masri, K., & Chu, D. (2019). Catch and release? NHL expansion draft endowment effects. Sport, Business and Management: An International Journal, 9(3), 301-312. doi:10.1108/SBM-04-2018-0028

Tingling, P., Masri, K., & Martell, M. (2011). Does order matter? an empirical analysis of NHL draft decisions. Sport, Business and Management: An International Journal, 1(2), 155-171.

Vincent, C., & Eastman, B. (2012). DOES PLAYER MOBILITY LEAD TO HIGHER EARNINGS? EVIDENCE FROM THE NHL. The American Economist, 57(1), 50-64. Retrieved from http://www.jstor.org.ezproxy.library.ubc.ca/stable/23240711