THE NHL DRAFT

Daniel Kim (47066162)

ECON 490: Seminar in Applied Economics

Under the supervision of Dr. Jonathan Graves

Abstract

With only a 0.16% chance of youth hockey players getting drafted into the NHL and only a 0.00025% chance of actually making it, the odds are definitely not favorable. However, players playing in a top junior league such as the Canadian Hockey League that comprise of 3 leagues — The Western Hockey league (WHL), The Ontario Hockey League (OHL) and The Quebec Major Junior League (QMJHL) — the transition rate to the NHL is one of the highest. However, it is almost impossible to reason why some players get drafted higher than others or why a player did not get picked. This is especially true to Defenseman since baseline statistics will not be able to accurately project their value. Luckily, advanced statistics is an up and coming field specializing in directly measuring player's values that is not possible with just baseline statistics. As a result, my research analyzes advanced statistics for both NHL drafted and undrafted CHL Defenseman to best predict draft outcome within their draft eligible years. With the use of the Heckman Selection Model as my methodology and exploiting multiple years of data, my results estimate that a unit change in Shots Per 60 (*S/60)* advanced statistic statistically significantly improves draft position by 18 positions. I also show that differing draft eligible years play a factor in how much draft position improves through the lens of advanced statistics. Near the end of my paper, I discuss any limitations to my results and solutions to it for future improvement.

Introduction

The National Hockey League (NHL) is a professional hockey league based in North America. There are 31 teams in total, with 7 teams spread across Canada and 31 teams in the United States. Every year, an NHL entry draft is conducted at the end of June and each 31 teams are allocated 7 picks per cycle. The order of the draft is based upon the most recent year standings, where the worst performing teams are given the highest pick and the best performing teams are given the last pick based on playoff performance and the regular season. In order to have integrity amongst the league and diminish strategic losses to gain the highest overall pick, a “lottery" system was put in place where the 15 teams that missed the playoffs are all given a chance to win the first overall pick. To create fairness, the teams that performed statistically worst based were given the highest percentage of winning the lottery in descending order.

One of the most intriguing aspects of the NHL Draft is determining why certain players were drafted and others not. A common conclusion casual fans think is that higher goals, assists and total points will yield a higher draft position (Canucks Army, 2015). However, this is not likely the case. By evaluating baseline statistics such as goals, assists, penalty minutes or total games played by a CHL player will at times seem random. In other words, just looking at a scoresheet doesn’t give the full picture (Schuckers, 2016). For example, arguably two of the best forwards in Brayden point and Nikita Kucherov of the Tampa Bay Lightning were drafted 74th overall in the 3rd round and 58th overall in the second round, respectively. Despite having very similar or even better baseline statistics to those drafted ahead, which leads to the question of why? Fyffe (2011) claims that these types of scenarios are common and frequent in almost all NHL Drafts. Hence, I find it appropriate to find a solution to this phenomenon and tackle the mystery behind the NHL Draft.

For the purpose of my research, I will be analyzing players in the Canadian Hockey League (CHL) which consist of 3 leagues, The Western Hockey League (WHL), The Ontario Hockey League (OHL) and The Quebec Major Junior League (QMJHL). The main motivation of my research is to find a more precise predictive measure of a player’s draft outcome that goes beyond baseline statistics such as goals, assists, penalty minutes and total games played. In particular, I will be focusing on CHL draft eligible prospects that predominantly play defense. Since

Defenseman do not contribute to the game by scoring goals, rather they aim to engage in more defensive strategies and contribute more to the team performance, baseline statistics won't will not tell the whole story. 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). Thus, I will be utilizing “advanced statistics". Hence, my research will be analyzing advanced statistics for both NHL drafted and undrafted CHL Defenseman to best predict draft outcomes within their draft eligible years.

The methodology I employ in my research is the Heckman Selection model for its ability to reduce sample selection bias and provide the most accurate representation in draft outcomes for CHL Defenseman. The way this model reduces selection bias is by accounting for all NHL drafted and undrafted CHL Defenseman across all years within my data set, but this would lead to the question of what makes this model more attractive than others?

Take an OLS regression as an example. The reason why I explicitly disregard an OLS regression for this project is because my main motivation is to distinguish drafted and undrafted CHL Defenseman with respect to advanced statistics. The issue is that an OLS regression will only account for drafted CHL Defenseman in my data set which means that it excludes the rest of the undrafted players. This would have been a huge limitation to my results, since there would have been a clear selection bias and ultimately defeat the purpose of my research. I find it fitting to utilize the Heckman Selection Model because it will give the most accurate measure of advanced statistics on draft position for all CHL Defensemen.

The way this model reduce selection bias is that it will give both NHL drafted and undrafted CHL Defenseman an arbitrary draft number. With this arbitrary draft number on all CHL Defenseman, the regression will be able to differentiate which advanced statistic can best signify draft outcome giving me a clear indication of why certain CHL Defenseman were drafted and tell me why the undrafted Defenseman did not get picked. An example of how to interpret the coefficients is when a player increases their Offensive Catalyst % by 1 unit, their draft position will either improve by an X amount of positions or decrease by an X amount. If the number comes out significant, I will be able to use this information to advise CHL Defensemen that if they improve their offensive catalyst %, they will become more desirable and increase their chances of being selected in their NHL Draft.

I ran 3 separate Heckman Models where the variable of focus is Beta 1 which are Shots Per 60 (*S/60)*, Even Strength Goals For Per 60 (*ESGF60)* and their Estimated Time On Ice Per Games Played (*ETOIGP)*. I also added control variables for each Heckman Model for the purpose of capturing potential noise and a dummy variable to see if different draft eligible years make a difference on draft position with respect to advanced stats.

CKM Sports Management provided player-by-player data on performance statistics and affiliated NHL (which NHL has their rights) attributes of all the players from years 2015 to 2019 that were eligible for the 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 – draft eligibility year (calculated from birth year), position on ice, junior league, and team changes. As my research requires more than baseline statistics, I was able to obtain player-by-player data on performance statistics that included advanced statistics through a lengthy independent research process. The first sources came from a Youtuber's website that dedicates its videos in hockey prospect analysis named Scouching. The second source is from an independent website that collects and presents advanced statistics of draft eligible prospects from years 2015 to 2019. I aggregated all three sources to evaluate consistency and accuracy in data, which came out to be legitimate. I will be mainly emphasizing the advanced statistic portion throughout the paper as I strongly assume this is what will cut right through the vagueness of why certain CHL Defensemen do and don't get drafted. Specific advanced statistics were dropped to provide value and relevance since certain advanced statistics did not give any value to the big picture of my research.

By exploiting my enlarged data set and with the help of the Heckman Selection Model, I found that Shots Per 60 (*S/60)* came out statistically significant, improving draft position by 18 position. This extended towards 3rd year draft eligible CHL Defenseman where players that improve their Shots Per 60 (*S/60)* advanced statistics astronomically improve their draft position by 76 positions. The one downside to my main results was that Even Strength Goals For Per 60 (*ESGF60)* and Estimated Time On Ice Per Game (*ETOIGP)* did not come out significant, despite showing improvement in draft position. To make sure my main coefficients are robust, I conduct 3 separate sensitivity checks per advanced statistics I analyzed. Furthermore, I also check the average squared residual for both the Heckman Selection Models and Sensitivity Analysis. With regards to the sensitivity analysis, all three advanced statistics produced very similar results, but with Shots Per 60 (*S/60)* actually improving in strength. The average squared residuals also demonstrated promise as my squared error residual decreased making my results legitimate. Yet, there are two limitations I discuss later in my paper that need consideration. The two limitations I disclose that would put my results into question are the obvious omitted variables and the data manipulation I operated in order to work within the constraints of the Heckman Selection Model. Near the end of this paper, I suggest ways to solve my limitations for future researchers.

This paper will be divided into sections. Section I will give background information of hockey advanced statistics and increasing importance to the hockey community. Section Ⅱ will describe the data employed. Section Ⅲ will give a detailed overview of the methodology. In Section Ⅳ, I will discuss the impact of advanced statistics and its significance towards the NHL Draft. Section Ⅴ will present any implications and Section Ⅵ will be the concluding remarks of this paper.

Ⅰ. Background

Ice hockey is played between two teams that consist of 6 players on the ice per team. These 6 players are divided into positions: 3 forwards, 2 defensemen and 1 goalie. However, the grouping of the players is due to change through the course of the game depending on the team's coaching strategy. The object is to propel a rubber disk, also known as the "puck", past a goal line and into a net guarded by a goaltender or "goalie" (Encyclopaedia Britannica, 2020). One distinguishing factor about ice hockey is its speed and its physical contact. Players on each team are allowed to "hit" their opponent with their shoulders as long as the opposing player has control of the puck or within 1.5 seconds after the puck has departed from the opposing player's control. Ice hockey has become one of the most popular international sports as it is one of the more prominent Olympic sporting events. With more than a million registered players performing regularly in leagues across all ages, it is perhaps Canada's more popular game (Encyclopaedia Britannica, 2020).

Many players around the globe aspire to play in the NHL since it is regarded as the premier professional hockey league consisting of the best players in the world. To play in the NHL, there are specific levels each player must compete in. These levels are usually divided by cohorts and by skill level, however, this varies across countries. Per Hockey Canada (2001), young aspiring hockey players go through each age cohort level accordingly: Pre-Novice (age ≤ 6), Novice (age 7-8), Atom (age 9-10), Pee-Wee (age 11-12), Bantam (age 13-14), Midget (age 15-17). Once a player is within the "Midget" cohort (age 15-17), these players are scouted more frequently by specific NHL teams in preparation for the NHL draft. However, due to stiff competition, excess number of players and limited roster spots in the NHL, the odds of becoming an NHL player is a very slim 0.00025% and it gets more difficult with age (Kalchman, 2003).

Players are eligible to be selected once they are in the year of turning 18 years old. Once a player is selected by their respective team, the team has the sole rights to the drafted player for roughly 3 years, during which time they may sign an entry-level contract. There are at most 217 drafted players in a single year and thus, thousands of other draft eligible players end up going undrafted. However, draft eligibility extends for 3 years between ages 18 to 20 years of age giving an undrafted player another opportunity to become a draft pick. However, the odds are not in favor for undrafted prospects as 90% of today's NHL is made up of drafted players (Nandakumar, 2017)

A. Literature Review

In recent years, after years of the NHL draft suffering from unchallenged bias and extensively relying on subjective "eye tests", hockey analytics has begun to pose new questions, crunch data and dive deep into inefficiencies in the scouting and drafting process. Analysts have been tackling numbers, seeking to uncover overrated and underrated draft prospects using innovative analytical techniques, advanced statistics or even through the intelligent application of conventional stats (Seppa, Schuckers and Rovito, 2016).

A popular model employed as a tool for the NHL Draft is the Bayesian Draft prediction model inspired by Brian Burke work on the NFL Draft prediction model (Hockey-Graphs, 2016). This model initially makes a "first guess" as to where a player will be selected and becomes progressively refined as new information is added. Lawrence (2015) and Weissbock (2015) developed a "cohort-based" model (PCS) which looked to predict NHL performance based on junior performance of players with similar characteristics (age, scoring rate, height), while adjusting for league strengths using NHL equivalencies. The idea behind PCS is that analysts can take a player and generate a list of comparable giving them an estimate of the likelihood of players becoming an NHL-er (Canucks Army, 2015). Tingling, Marsi and Martell (2011) show that beyond the first three rounds of the NHL Draft, there is no significant difference in drafted players' success rates. Sprigings (2016) created a draft probability model for several highly rated players in the 2016 NHL Draft to inform teams' decision to trade picks in the hopes of acquiring a specific prospect. Seppa, Schuckers and Rovito (2016) found that advances statistics utilized such as even-strength goal scoring rate (ESG/60) and even-strength primary assists rate (ESA1/60) yield improved performance of junior league stats in evaluating best pro-prospects, but shows that text-mined scouting data is more predictive on its own than the performance-based analytics (advanced statistics). Seppa, Schuckers and Rovito (2016) concluded draft day decisions should rest in the hands of front office personnel who understand and can incorporate all the insights of performance based and scouting based analytics.

Since most of the studies do not specifically explore draft outcomes relative to advanced statistics, this leads to my research. The studies mentioned above in this section do explore and employ advanced statistics, but primarily to uncover future projections of drafted prospects becoming NHL players. The purpose of this research is to explore both drafted and undrafted players with respect to advanced analytics in hope to predict draft outcomes for CHL defenseman. Due to the lack of research conducted in this particular topic, it is necessary to pursue this research question with consideration of other studies.

II. Data

The performance statistics amongst all CHL players were accumulated through 3 main sources. From these 3 sources, the total amount of observations amounted to roughly 6900. The first dataset collected consists of baseline statistics (total goals, assists, points, penalty minutes) of drafted CHL players and their affiliated NHL team. This data set ranges from 2015 to 2019, however, year 2018 is pending approval. This CHL-NHL data set was obtained through CKM Sports Management, a sports agency firm working closely with this project. The two other main data sets are from third party sources.

The first data set obtained through a third party originated from a hockey analytics fanatic called "Scouch" (https://www.scouching.ca/data_archive.html), where this person uploads raw prospect data from multiple leagues around the globe. This includes my research specific setting, the CHL and other junior leagues such as the Super-Elite League in Sweden and the NCAA in the United States. However, due to relevance, I will be dropping players outside of the CHL. What makes this data set very attractive due to its "Advanced Statistics" content that goes beyond baseline statistics and observations of undrafted players. This will help my research since it will give my research more information to work with in my model. The "Scouching" data is available from years 2017 to 2019.

The second third party data set obtained consisted of all drafted players from years 2015 to 2019 and was obtained through an online hockey data source called "Pick224" (https://pick224.com/). This data set also contains advanced statistics for drafted and undrafted players. All three datasets contained Forwards, Defenseman and Goalie. Since my research will be specified to CHL Defenseman, I dropped observations of forwards and goalies. This data also included European 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. Through the supervision of Dr. Jonathan Graves, the third-party datasets have been granted approval for use in this research study.

A. Data Setup and Transformation

All three datasets were imported individually then merged into one superset. This way, I will be able analyze the advanced statistics aggregated across all the years and have all relevant observations in hand. Due to the wide range of players in different leagues, players that played outside of the CHL in the Scouching and Pick 224 data were dropped due to a lack of statistics available or skewed values from playing limited games. Furthermore, I have dropped all CHL players with the role “Forwards” since my research is evaluating Defenseman. Dummy variables were created for qualitative variables such as “Eligible Draft Year” (EDY). For simplicity, I denoted “Eligible Draft Year” as “EligYear”.

$$Eligible\ Draft\ Year\ (EligYear) = 1, if\ first\ eligible\ draft\ year$$

$$Eligible\ Draft\ Year\ (EligYear + 1) = 2, if\ second\ eligible\ draft\ year$$

$$Eligible\ Draft\ Year\ (EligYear + 2) = 3, if\ third\ eligible\ draft\ year$$

The properties of the three aggregated advanced statistics are identified as quantitative, qualitative and observational.

Advanced statistic variables I will be using in the regression include Time On Ice (*TOI),* Estimated Time On Ice Per Game (*ETOIGP*), Shots Per 60 minutes (*S/60*), Even Strength Goals For Per 60 minutes (*ESGF60*), Even Strength Goals Against Per 60 minutes (*ESGA60),* Involvement Percentage (*Inv%)* and Offensive Catalyst Percentage (*OFFCAT%*).

TOI refers to how long a player has played in a single game. *ETOIGP* estimates how much playing time or ice time a player will receive per game. *S/60* is an estimate of the amount of shots a player makes per 60 minutes — 60-minute regular game; excluding playoffs.

$$\frac{S}{60} = [Aggregated\ total\ shots\ taken\ in\ a\ game]/60$$

ESGF60 measures a player’s ability to increase even strength offense by over a goal and a half per hour. *ESGF%* is calculated by aggregating total even strength goals (ESG) divided by time on ice (*TOI*) times games played per 60 minutes.

$$ESGF60 = \frac{ESG}{\left[TOI \times \left(\frac{GP}{60}\right)\right]}$$

ESGA% is relates to *ESGA%*, but in opposite terms. This advanced stat evaluates a player's ability to decrease even strength goals against by up to a goal per game. Both *ESGF%* and *ESGA%* complement each other as the equation above can produce either of these statistics, depending on what is being tracked; goals scored (GF) or goals against (GA). *Inv%* evaluates how offensively involved a player is a given game. This metric is defined as player points per game divided by their respective team's goals per game.

$$Inv\% = \frac{[(Player's\ points)/GP]}{[Team's\ goals/GP]}$$

The last advanced statistics I will be using is *OFFCAT%*. This statistic considers how much of an effect a player has on their ability to drive *ESGF%* which can be evaluated to see offensive competence (Couching, 2018). The mathematical equation is as follows:

$$[ESGF60\ or\ ESGA60\ or\ ESGF\%Rel] - 1 = Offensive\ Catalyst\ \%$$

One adjustment I have made in my data set after appending years 2015 to 2019, I generated a second variable doubling each of the advanced statistics — Estimated Time On Ice Per Game (*ETOIGP*), Shots Per 60 minutes (*S/60*), Even Strength Goals For Per 60 minutes (*ESGF60*), Even Strength Goals Against Per 60 minutes (*ESGA60),* Involvement Percentage (*Inv%)* and Offensive Catalyst Percentage (*OFFCAT%*) — and created additional dummy variables since not all CHL Defenseman had these statistics available. For example, some players in my data set have statistics for Shots Per 60 minutes (*S/60*) and Estimated Time On Ice Per Game (*ETOIGP*), but not Offensive Catalyst Percentage (*OFFCAT%*). The additional variables I produced in conjunction with my original data set, the Heckman Selection Model will account for all observations despite lacking one or two missing advanced statistics from the appended data set, thus increasing total observations, n.

The reason for putting such emphasis on "goals for" (GF) is because it quantifies a player's capabilities of producing offense and driving play. The NHL is becoming more of a faster paced and offensively driven league (Sports Illustrated, 2015), especially for Defenseman with the likes of Quinn Hughes and Cale Makar making an incredible impact in the game. Hence, these advanced stats can potentially be a determining factor to a CHL defender's draft outcome.

Per Couch (2018), these are the variables deemed important in analyzing a player's offensive capability. As these metrics are universal amongst Forward and Defenseman, I will be accepting these variables to determine a CHL Defenseman draft outcome. Below, Table Ⅰ is showing the summary statistics for this research project.

Table 1: Summary Statistics

	Mean	St.Dev	min	max	Median
Inv%	.24	.089	.055	.43	.24
S/60	7.154	3.384	0	13.351	7.75
GF%Rel	5.107	8.22	-10.75	24.88	4.71
TOI	14.423	1.928	7.97	18.39	14.13
ESGF60	3.479	.778	1.842	5.418	3.49
ESGA60	2.864	.839	1.449	5.498	2.717
ESGF60rel	1.229	.713	-.301	2.517	1.256
ESGA60rel	.625	.462	-.233	2.13	.597
OffCat%	.605	.389	-.09	1.585	.59
ETOIGP	16.068	3.54	8.1	28.51	15.895

Ⅲ. Methodology

For this research, I will be using a two-step statistical model called the Heckman Selection Model. I will be incorporating the advanced hockey statistics as variables mentioned in the data section obtained through third party sources. Since my research is focusing on Defenseman, I will be employing most of the advanced stats, but only the ones the advanced statistics community deem worthy and impactful for Defenseman. It is also critical to mention that I will also be incorporating control variables in this regression model to capture potential noise. Eligible draft year (EligYear = 1, 2, 3) will be included into the regression as a dummy variable to see if there are any changes on draft position in different years of eligibility.

Below is the preliminary regression specification to estimates the draft position based on advanced statistics:

$$Y_{\mathrm{i}} = B_0 + B_1\left(\frac{S}{60}\right) + B_2 ETOIGP + B_3 ESGF60 + B_4 ESGA60 + B_5 Inv\%$$
$$+ B_6 Offensive\ Catalyst\ \% + B_7 EligYear\ (1,2,3) + \varepsilon$$

Where Y_i is "*Draft Position*", for every player i, *S/60* evaluates the amount of shots a player makes per game, *ETOIGP* estimates the amount of playing time a player will receive per game, *ESGF%* measures a player's ability to increase even strength offense by over a goal and a half per hour (60 minutes), *ESGA%* evaluates a player's ability to decrease even strength goals against by up to a goal per game, *Inv%* evaluates how offensively involved a player is a given game, *OFFCAT%* is how much of an effect a player has on their ability to drive offence and *EligYear* presents specific draft year eligibility as a dummy variable.

Through an OLS regression, Draft Position is only going to account for CHL Defenseman who got drafted by an NHL team and not for CHL Defenseman that didn't get picked. Thus, there will be bias towards Drafted players and will not be an adequate method in distinguishing both Drafted and Undrafted CHL Defenseman. Hence, this regression alone will lead to a sample selection problem since not all observations are captured, but rather a subset (Bushway, Johnson and Slocum, 2007). Which leads to why I will be utilizing the Heckman Selection Model.

The Heckman Selection Model will account for variables with missing values, reduce selection bias and give a more accurate coefficient (Puhani, 2002). Furthermore, employing this model will allow me to accurately distinguish between drafted and undrafted CHL Defenseman in the NHL Entry Draft. With the Heckman Selection Model, a 2-step equation will address this selection bias:

$$(1)\ Draft\ position_{I}^{*} = XB' + \varepsilon_{i}$$

$$(2)\ Draft\ outcome_{I}^{*} = Z\gamma' + \varepsilon_{i}$$

This means Draft position = Draft position$_i^*$ IFF Draft outcome$_i^*$ > 0 and Draft position = . IFF Draft outcome$_i^* \leq 0$.

The first stage of the Heckman Selection Model uses a probit model (Equation 2) to estimate the probability of CHL Defenseman getting drafted into the NHL. The second stage uses OLS (Equation 1) to predict the ultimate dependent variable estimate (Certo, Busenbark, Woo, Semadeni, 2015)

For my research, instead of inputting all the advanced statistics into one model as shown above, I will be analyzing 3 separate advanced statistics individually by regressing 3 independent Heckman Selection Models. The 3 advanced stats I will be using as a proxy to analyze its effect on draft position for CHL Defenseman are Shots Per 60 (*S/60)*, Even Strength Goals For Per 60 (*ESGF60)* and Estimated Time On Ice Per Game (*ETOIGP)*.

$$Draft\ position_{i}^{*} = B_0 + B_1\left(\frac{S}{60}\right) + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\%$$
$$+ B_5 EligYear + \varepsilon$$

$$Draft\ position_{i}^{*} = B_0 + B_1 ESGF60 + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\%$$
$$+ B_5 ELigYear + \varepsilon$$

$$Draft\ position_{i}^{*} = B_0 + B_1 ETOIGP + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\%$$
$$+ B_5 EligYear + \varepsilon$$

Furthermore, I will also accompany the 3 explanatory variables with the rest of the advanced statistics indicated in the data section and shown in the preliminary regression specification as control variables for the purpose of capturing noise. These control variables include Offensive Catalyst Percentage (*OffCat%)*, Even Strength Goals Against (*ESGA60)* and Involvement

Percentage (*Inv%)*. These three advanced stats will ultimately allow me to clearly see if my hypothesis of having a greater "goals for " or offensively ability is a significant contributing factor in selecting CHL Defenseman higher in the NHL Draft. In the next section, I will be showing my results through 2 perspectives: OLS regression and the Heckman Selection Models.

IV. Results

A. OLS Regression

Table 2 evaluates the preliminary OLS linear regression for draft position on hockey advanced statistics. In column (1), the results indicate that a higher Shots Per 60 (*S/60)* correlates to having a higher draft position in the NHL draft. This can be interpreted as: a 1 unit increase in Shot Per 60 (*S/60)* metric, draft position improved by 5 positions for CHL Defenseman. Despite the improved draft position, it is not statistically significant.

Table 2: Preliminary OLS Results

Draft position	Coef.	t-value	p-value	[95% Conf	Interval]
Shots per 60	-5.526 (8.163)	-0.68	0.500	-21.746	10.695
Even Strength Goals For 60	-21.455 (26.268)	-0.82	0.416	-73.649	30.739
Estimated Time On Ice GP	-10.018 (13.592)	-0.74	0.463	-37.026	16.989
Offensive Catalyst %	-27.016 (54.094)	-0.50	0.619	-134.504	80.473
Involvement %	-116.021 (299.328)	-0.39	0.699	-710.779	478.737
Even Strength Goals Against 60	9.961 (20.868)	0.48	0.634	-31.502	51.425
Eligible Year = 2	45.228** (17.261)	2.62	0.010	10.930	79.526
Eligible Year =	28.363	0.98	0.330	-29.122	85.848

3	(28.931)		
Mean dependent var	97.812	SD dependent var	58.897
R-squared	0.192	Number of obs	101.000
F-test	1.922	Prob > F	0.047

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

I would like to mention that the negative coefficient is correct because in the NHL Entry Draft, a higher draft position is reflected by a lower numerical value. For example, having a draft position of "3rd" is presented as being a "higher draft pick" or having a higher "draft position" rather than a player being picked "10th". Thus, the negative relationship between advanced statistics and draft position is desired in my research. For even Strength Goals For Per 60 minutes *(ESGF60)*, it shows that with a 1 unit increase, it will result in a higher draft position by 21 positions. However, it is not statistically significant. Estimated Time On Ice Per Game *(ETOIPG)* also yield improvement in draft position by 10 positions with a 1 unit increase, but it is not statistically significant.

The reason why I produce results with a OLS regression is because I wanted to make sure my concerns with OLS were honest. Per table 2, these results would not be able to tell me what makes drafted CHL Defenseman more attractive than those not drafted, which is the main reason why an OLS linear regression is not a viable option in my research analysis. One reason why the three advanced statistics in focus are not be showing significance is likely due to a selection bias where the linear regression is only accounting for drafted CHL Defenseman or it could be due to a very low number of observations evaluated. This gives me the empirical proof that an OLS is not a good solution for my research and that a Heckman Selection Model is the best tactic to continue with for my research purposes.

B. Heckman Selection Model

In this section, I would like to reiterate the 3 advanced statistics I will be analyzing individually. These 3 advanced statistics include Shots Per 60 (*S/60),* Even Strength Goals For Per 60 (*EVGF60)* and Estimated Time On Ice Per Game (*ETOIGP).* Each of the three regressions will be complemented with control variables for the purpose of capturing potential noise. These control variables include Even Strength Goals Against Per 60 minutes (*ESGA60),* Involvement Percentage (*Inv%)* and Offensive Catalyst Percentage (*OFFCAT%*). Furthermore, the addition of Eligible Draft Year, denoted as "*EligYear",* will provide me additional context of how draft positions will vary depending on which year a player's draft eligible year is.

The first Heckman Selection Model regression I will be employing investigates Shots per 60 (*S/60).* The regression is as follows:

$$Draft\ position_i^* = B_0 + B_1\left(\frac{S}{60}\right) + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\% + B_5 EligYear + \varepsilon$$

In table 3, row (1) supports the notion that Shots Per 60 (*S/60)* does improve draft position. More specifically, with a 1 unit increase in Shorts Per 60 *(S/60)*, it statistically significantly increases a CHL Defenseman's draft position by 18 positions at the 10% level. Row (6) indicates that a CHL Defenseman in his third year of draft eligibility represented by a dummy (*EligYear = 3)*, their draft position improves by 76 positions and it is statistically significant at the 10% level. This tells me that shooting more in games will in fact improve a CHL Defenseman's draft position, which is exciting to see because this information can be simply and directly translated to upcoming CHL Defenseman.

Table 3: Heckman Selection Model Regression 1

Draft position	Coef.	t-value	p-value	[95% Conf	Interval]
Shots/60	-18.006* (9.863)	-1.83	0.068	-37.338	1.326
Offensive Catalyst%	-29.097 (67.894)	-0.43	0.668	-162.168	103.974
Even Strength Goals Against per 60	-13.977 (26.855)	-0.52	0.603	-66.612	38.658
Involvement%	379.766 (331.363)	1.15	0.252	-269.692	1029.225

Eligible Year = 2	-28.064 (26.235)	-1.07	0.285	-79.483	23.355
Eligible Year = 3	-75.637* (40.211)	-1.88	0.060	-154.448	3.175
Mean dependent var	97.812	SD dependent var			58.897
Number of obs	1147.000	Chi-square			7.010

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

The second Heckman Selection Model regression that is being investigates is Even Strength Goals For Per 60 (*EVGF60).* The regression is as follows:

$$Draft\ position_i^* = B_0 + B_1 ESGF60 + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\% + B_5 ELigYear + \varepsilon$$

Refer to table 4 where it presents the impact on draft position when increasing a CHL Defenseman's Even Strength Goals For Per 60 (*ESGF60)* advanced statistic. In row (1), the coefficient does represent improvement in draft position by 6 positions, with a 1 unit increase in Even Strength Goals For Per 60 (*ESGF60)*, however it is not statistically significant. Unfortunately, row (1) indicates that Even Strength Goals For Per 60 (*ESGF60)* is in fact not a strong predictive performance statistic on draft position for CHL Defenseman, but it should not be entirely disregarded. The reason is because it intuitively makes sense that a CHL Defenseman that generates more goals per 60 minutes equate to a higher draft position. The caveat to this is that it can't be said as certain.

Row (6) shows that the dummy variable for 3rd year draft eligible CHL Defenseman (*EligYear = 3)* shows improvement in draft position by 83 positions and this is statistically significant at the 5% level. In other words, players in their 3rd of draft eligibility having higher Even Strength Goals For Per 60 (*ESGF60)* does in fact improve their draft position. This is interesting to see because 3rd year draft eligible CHL player's probability of getting drafted exponentially drops compared to players in their first or even their second draft eligible years (Fyffe, 2009). This to me is very encouraging, especially for over aged CHL Defenseman that were passed over in multiple drafts since it will give "over agers" hope that there is a measureable performance statistics they can focus on knowing the fact that it will improve their draft outcome and take it as if it is their last chance.

Table 4: Heckman Selection Model Regression 2

Draft position	Coef.	.	t-value	p-value	[95% Conf	Interval]	
Even Strength Goals For 60	-5.839 (36.398)		-0.16	0.873	-77.177	65.499	
Offensive Catalyst %	-20.820 (75.970)		-0.27	0.784	-169.719	128.079	
Even Strength Goals Against 60	-6.529 (26.153)		-0.25	0.803	-57.787	44.730	
Involvement %	173.423 (341.074)		0.51	0.611	-495.070	841.917	
Eligible Year = 2	-31.173 (26.214)		-1.19	0.234	-82.552	20.206	
Eligible Year = 3	-83.353** (40.246)		-2.07	0.038	-162.233	-4.473	**
Mean dependent var		97.812	SD dependent var			58.897	
Number of obs		1147.000	Chi-square			5.867	

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

The third Heckman Selection Model regression evaluates Estimated Time On Ice Per Game (*ETOIGP).* The regression is as follows:

$$Draft\ position_i^* = B_0 + B_1 ETOIGP + B_2 OffCat\% + B_3 ESGA60 + B_4 Inv\% + B_5 EligYear + \varepsilon$$

Refer to table 5 where it presents the impact on draft position when CHL Defenseman increase their Estimated Time On Ice Per Game (*ETOIGP)* advanced statistic. In other words, does increasing a CHL Defenseman's playing time increase or decrease their draft position during their draft eligible years? Row (1) shows that a 1 unit increase in Estimated Time On Ice Per Game (*ETOIGP),* draft position improves by 5 positions, but it is not statistically significant.

Similar to the second Heckman Selection Model, it can still be informative to CHL Defenseman since the intuition of the coefficient is correct. The reason I say this is because it is widely known that the best players, especially for Defenseman, tend to play the most minutes in a game due to having a lower number of defenders dressed in a game comparative to forwards (Hockey Canada, 2001). With the CHL arguably being the best feeder league into the NHL and league's best Defenders yield higher draft position, this information can be utilized to inform CHL Defenseman that getting the as much ice time per game is one of the best possible way to display their offensive abilities to scouts. One caveat to this is that one can’t exclusively claim a higher Estimated Time On Ice Per Game (*ETOIGP)* will definitely equate to a higher draft position because the coefficient is not statistically significant, but it can definitely be employed as a complementary piece.

Table 5: Heckman Selection Model Regression 3

Draft position	Coef.	t-value	p-value	[95% Conf	Interval]
Estimated Time On Ice Per Game	-5.435 (12.558)	-0.43	0.665	-30.049	19.178
Offensive Catalyst%	-32.651 (69.178)	-0.47	0.637	-168.237	102.935
Even Strength Goals Against 60	0.111 (26.126)	0.00	0.997	-51.096	51.317
Involvement%	289.964 (332.008)	0.90	0.368	-341.160	921.088
Eligible Year = 2	-32.258 (25.472)	-1.27	0.205	-82.183	17.666
Eligible Year = 3	-85.654** (39.942)	-2.14	0.032	-163.940	-7.368
Mean dependent var		97.812	SD dependent var		58.897
Number of obs		1147.000	Chi-square		6.607

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

Row (6) indicates that CHL Defenseman in his 3rd year of draft eligibility (*EligYear = 3)* will improve their draft position by 86 positions with a 1 unit increase in Estimated Time On Ice Per Game (*ETOIGP).* I interpret this coefficient that playing time is critical, especially for undrafted 3rd year CHL Defenseman and that I would advise all players to go to a team that would give them the largest amount of playing time since playing time can make or break a hockey career (Boost Hockey, 2019)

V. Discussion

In this section, I will be separating it into 3 components. The first section is where I will be discussing my robustness check. In the second section, I will parse the results on how family advisors can interpret my results and what course of action would best suit for CHL Defenseman. Lastly, I will use the third section to discuss potential limitations within my analysis.

A. Robustness check

For my Robustness check, I will be employing a sensitivity analysis. According to Gurra and Neuberger (2010), the purpose of the sensitivity analysis is to examine how changes in the assumptions of an economic model affects its predictions. Ultimately, it is a valuable modeling tool because it may provide information on the robustness of my model's predictions.

In my research, I will be replacing my control variables with a completely different set of hockey advanced statistics. These advanced statistics are Goals For Relative to Team (*GF%Rel),* Even Strength Goals For Per 60 Relative to Team (*ESGF60Rel)* and Even Strength Goals Against Per 60 Relative to Team (*ESGA60Rel).* These advanced statistics will be incorporated into the Heckman Selection Model and be evaluated with the 3 explanatory variables — Shots Per 60 (*S/60),* Even Strength Goals For Per 60 (*ESGF60),* Estimated Time On Ice Per Game (*ETOIGP) —* but through a team's perspective. The motivation behind utilizing team specific offensive related advanced statistics is to analyze if the sensitivity model will yield similar results when considering team characteristics. The three sensitivity check regressions look as follows:

$$Draft\ position_{\mathrm{i}}^{*} = B_0 + B_1\left(\frac{S}{60}\right) + B_2 GF\%Rel + B_3 ESGF60Rel + B_4 ESGA60Rel$$

$$+ B_5 EligYear + \varepsilon \quad (1)$$

$$Draft\ position_i^* = B_0 + B_1 ESGF60 + B_2 GF\%Rel + B_3 ESGF60Rel + B_4 ESGA60Rel$$
$$+ B_5 EligYear + \varepsilon \quad (2)$$

$$Draft\ position_i^* = B_0 + B_1 ETOIGP + B_2 GF\%Rel + B_3 ESGF60Rel + B_4 ESGA60Rel$$
$$+ B_5 EligYear + \varepsilon \quad (3)$$

Table 6 presents the results of the sensitivity analysis. Each column in table 6 presents each of the sensitivity check regressions where columns (1) represents Shots Per 60 (*S/60),* (2) represents Even Strength Goals For Per 60 (*ESGF60)* and (3) represents Estimated Time On Ice Per Game (*ETOIGP).* It can be seen through the 3 columns that the sensitivity coefficients and my main Heckman Selection Model regression coefficients are in fact similar.

Table 6: Sensitivity check

Draft position	(1) Sen1	(2) Sen2	(3) Sen3
Shots/60	-15.811** - (7.520)		
Even Strength Goals For 60		-14.307 (32.278)	
Estimated Time On Ice Per Game			-9.267 (9.356)
Goals For % Relative	-3.250 (3.133)	-3.052 (3.128)	-2.939 (3.024)

Even Strength Goals For 60 Relative	44.731 (46.545)	42.926 (52.940)	53.152 (47.728)
Even Strength Goals Against 60 Relative	6.613 (46.090)	-18.147 (46.977)	-5.650 (50.125)
Eligible Year = 2	-6.875 (18.936)	-4.670 (19.156)	-12.958 (19.467)
Eligible Year = 3	-46.884 (29.274)	-47.582 (29.687)	-59.551** (29.502)

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

In column (1), the results between the first sensitivity check and my main results appear similar since it shows that a 1 unit increase in Shots Per 60 (*S/60)* statistically significantly improves draft position by 16 positions in the sensitivity analysis. In Column (2), even Strength Goals For Per 60 (*ESGF60)* has a coefficients that shows improvement in draft position, but is not statistically significant which is similar to my main analysis. In Column (3), Estimated Time On Ice Per Game (*ETOIGP)* has a coefficients that indicates improvement in draft position, but it is not statistically significant which is also similar to my main results. Furthermore, my p-value for Shots Per 60 (*S/60)* improved from being 0.068 where it is significant at the 10% level to 0.035 being significant at the 5% level. However, there was little to no change between my sensitivity check and my main results for Even Strength Goals For Per 60 (*ESGF60)* and Estimated Time On Ice Per Game (*ETOIGP)* with regards to their p-value. Refer to table 7 for a direct comparison of the coefficients between my sensitivity analysis and my main analysis.

Table 8 presents a fit test where I am checking the averaged squared residuals for the main coefficients — Shorts Per 60 (*S/60),* Even Strength Goals For Per 60 (*ESGF60),* Estimated Time On Ice (*ETOIGP)* — between the main Heckman Selection Model and the Sensitivity analysis. The purpose of this test is to double reference my results to make sure they are strong. The way a fit test can declare my results as strong is by checking if my averaged squared residual decreases. This can be seen through the standard error column denoted as “Std.Err”. The rows represent the comparison of the Heckman Selection Models and the respective sensitivity analysis per advanced statistic.

Row (1) shows that the averaged squared residual decreases in the sensitivity check, meaning that my sensitivity analysis regarding Shots Per 60 (*S/60)* is stronger. This means my sensitivity check with respect to Shots Per 60 (S/60) is as or even more accurate than my main coefficient. Regardless, since both coefficients are statistically significant, I declare Shots Per 60 (*S/60)* as a significant advanced statistic that does indeed play a critical role in determining draft outcome. Row (2) and (3) present Even Strength Goals For Per 60 (*ESGF60*) and Estimated Time On Ice Per Game (*ETOIGP*) respectively. It shows that the averaged squared error residuals for both Even Strength Goals For Per 60 (*ESGF60*) and Estimated Time On Ice Per Game (*ETOIGP*) is lower in the sensitivity analysis in comparison with the Heckman Selection Model. With this in mind, the results from the sensitivity analysis further complements my main results as it signifies that Even Strength Goals For Per 60 (*ESGF60)* and Estimated Time On Ice Per Game (*ETOIGP)* is not a good indicator in determining draft outcome for CHL Defenseman. Thus, I conclude that my results are robust and solid.

B. Interpretation

Family advisors, coaches or scouts may ask how to interpret my results to their clients. One important note that needs understanding is that advanced statistics is by no means a direct measure of how good a player is. The motivation behind this project is to help players increase their chances of becoming a NHL draft pick and fulfilling a lifelong dream. Family advisors, coaches, parents or scouts should not be using my results as a way to label a player as being a boom or bust, but use it as a measuring stick for refinement.

A way Shots Per 60 (*S/60)* can be translated to a CHL Defenseman is to simply tell a player to get as many shots directed at the net. Obviously, this is an over-simplification, but it does hold merit because as a player increases the sheer amount of shots in a game, the overall statistic for Shots Per 60 (*S/60)* will increase. A way a family advisor or a development coach can help catalyze this mindset of always shooting is first to show these players my results that getting the puck on net as much as possible does in fact make a big impact in getting drafted for Defenseman.

Furthermore, it can also be shown through video clips of their client in game action where the family advisor or coach can pin-point situations of how they can improve their shooting on net. Another approach is to dissect clips of previous NHL drafted CHL Defenseman or current NHL Defenseman on how they are able to create space allowing them to get more shots directed at the net since my data suggest that CHL Defenseman getting picked earlier do have higher Shots Per 60 (*S/60)* stats. These clips can then be further applied on the ice through skill sessions where they practice the studied clips.

C. Limitations

Despite my research being robust, there are two empirical limitations that need to be addressed. I will be discussing this matter in two sub-categorized sections. The first section will discuss how my data manipulation and second section will address omitted variable bias.

Section I

The main challenge my empirical research faces is the data manipulation I conducted to fit within the strict rules of the Heckman Selection Model. One of the challenges that came when obtaining my data from third party sources was that not every observation had advanced statistics available. For example, in my Pick224 data set from years 2015 to 2019, there were a handful of players per year with missing advanced statistics. This was prevalent with the Scouching data as well. Due to missing advanced stats, the Heckman Selection model completely neglected observations that did not have the advanced statistics. Thus, the total number of CHL Defenseman evaluated through the Heckman Selection model was consistently low which would immediately raise questions about my results. To fix this issue, I created an additional variable for each of the advanced statistics I analyzed where I indicated observations that did not have advanced statistics available in the data as blank. I then created dummy variables for each of the advanced statistics indicating as 1 for observations that have the advanced statistic and 0 otherwise in order to average out the effects for observations with a new value of 0. This way, the Heckman Selection model would include all CHL Defenseman in the model. The empirical issue is that there is still a level of uncertainty with the averaged-out effect since we won't know whether the missing data is systematically missing; in which case, my results would be biased. Due to this ambiguity and the unknown surrounding the legitimacy of my data manipulation to accommodate for the Heckman Selection Model, it is a limitation to my research.

Table 7: Direct comparison between Sensitivity Coefficient and Heckman Selection Model

Draft position	(1)		(2)		(3)	
	Sen1	Main1	Sen2	Main2	Sen3	Main3
Shots/60	-15.811**	-18.006*				
	(7.520)	(9.863)				
Even Strength Goals For 60			-14.307	-5.839		
			(32.278)	(36.398)		
Estimated Time On Ice Per Game					-9.267	-5.435
					(9.356)	(12.558)
Eligible Year = 2	-6.875	-28.064	-4.670	-31.173	-12.958	-32.258
	(18.936)	(26.235)	(19.156)	(26.214)	(19.467)	(25.472)
Eligible Year = 3	-46.884	-75.637*	-47.582	-83.353**	-59.551**	-85.654**
	(29.274)	(40.211)	(29.687)	(40.246)	(29.502)	(39.942)

**notes: Sen1 represents sensitivity check for S/60, Sen2 represents sensitivity check for ESGF60, Sens3 represents sensitivity check for ETOIGP*
**side-to-side comparison between main coefficient in Heckman Selection Model and the Sensitivity analsysis*
**** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Section Ⅱ

The second empirical limitation to my research is the clear omitted variables in my research. Baker and Logan (2007) identifies a range of environmental and genetic factors that can directly or indirectly influence a player's performance. With factors, such as cognition, genetic characteristics, socio-economic status, scouting evaluations excluded from my model, my coefficients may have resulted differently. For example, a hockey player's cognition is commonly referred to as "hockey I.Q" where is signifies how well a player can think the game. This can be in forms of how well a player handles pressure, game understanding, their positioning in varying situations, timing of their shots, their playmaking ability, pattern recognition and problem solving (The Blue Line Hockey I.Q Program, n.d.). The reason why omitting cognition is a limitation to my research is because all of the descriptors I just mentioned unequivocally relate to all player's advanced statistics I analyze because hockey I.Q would play a factor into offensive capabilities. By adding cognition into my model, it may change my results significantly to the point where my results may in fact be null or suggest otherwise.

With respect to genetic factors and socio-economic status, numerous NHL teams draft players, especially Defenseman not always entirely based on statistics, but based on how physically mature or slight the player is. The reason NHL teams do this is because they either want two things: 1) NHL teams are trying to "buy" potential. This means that NHL teams are drafting players strictly based on their hockey abilities and banking on them becoming a far better player down the line once they physically mature. 2) Some NHL teams also draft players based on how physically mature they are at the moment of being picked because they are in an immediate need for players to come in to fill a roster spot (Canucks Army, 2015). The issue with having genetic traits and socio-economic status omitted from my model is because my results may be positively bias because of it. Since strategic drafting with respect to genetic factors and socio-economic status is prevalent, my results may be null when accounting for these variables in my model. Further research in this area would be beneficial to see how these variables will impact the importance of advanced statistic on draft position.

Table 8: Fit-Test

	Coefficient	Std.Err.	[95%_Con	Interval]
Main (*S/60*)	24029.57	2121.266	19821.04	28238.1
Sens (*S/60*)	13292.35	1353.925	10606.2	15978.5
Main(*ESGF60*)	26448.44	2248.592	21987.3	30909.58
Sens(*ESGF60*)	13498.99	1363.234	10794.37	16203.61
Main(*ETOIGP*)	21963.51	1997.632	18000.26	25926.75
Sens(*ETOIGP*)	13940.92	1419.651	11124.37	16757.46

**note: Std.Err represents squared error residual in column (2)*

However, the issue with accounting for these omitted variables is that they are hard to quantify. Ultimately, cognition, genetic factors and social-economic status are objective measures making it hard to implement into a statistical model. One solution to this is for future researchers to utilize scouting files of draft eligible CHL Defenseman since these files will contain all information about each player (Brown, 2019). With the files in hand, a consensus of the level of hockey I.Q, the genetic make-up and labelled socio-economic status can be made. This would allow future researchers to include these variables as dummy variables into their model and see how much of an affect these variables will contribute towards advanced statistics with respect to draft outcome. This is also another limitation to my model is that I was not able to compile these scouting evaluations of all the CHL Defenseman in my data set. By incorporating scouting evaluations into my model, it could have legitimized my results further or could have re-directed my research into a different, but more accurate path with regards to advanced statistics. In the end, the whole desire for my research is to give as much informative knowledge to aspiring CHL defenseman and being able address these limitations will be invaluable.

VI. Conclusion

The NHL Draft for many years has been a long sought after mystery when it comes to reasoning why some players are drafted higher while other players drop or not get picked at all. This is especially true for Defenseman due to the nature of the position since the main objective for defenders is to prevent the puck from going in their team's net. However, with the NHL becoming more of a faster paced and offensively driven league, NHL teams have been putting a premium on offensively minded Defenseman when drafting (Sports Illustrated, 2015). Due to the vagueness of baseline statistics, this geared me towards advanced hockey statistics in hopes to find a concrete relationship with respect to draft outcome specifically for defenseman. As a result, my research analyzes advanced statistics for both NHL drafted and undrafted CHL Defensemen in order to best predict draft outcome within their draft eligible years.

I strictly studied CHL Defenseman in the Canadian Hockey League because this league is known to be one of the best to yield NHL draft picks (Canadian Hockey League, n.d.). The 3 advanced statistics I focused on are Shots Per 60 (*S/60*) where this stat estimates the amount of shots a player makes per 60 minutes, Even Strength Goals For Per 60 minutes (*ESGF60*) which estimates a player's ability to increase even strength offense by over a goal and a half per hour and Estimated Time On Ice Per Game (*ETOIGP*) measuring how much playing time a player will receive per game. The reason why I focused on these advanced statistics is because it best quantifies a player's capabilities of producing offense (Couch, 2018).

I found the amount of shots a CHL Defenseman makes per 60 minutes is a very important factor in distinguishing drafted and undrafted CHL Defenseman since my results statistically significantly indicate that higher positioned Defenseman picked have a higher Shots Per 60 (*S/60)* statistic. On the other hand, Even Strength Goals For Per 60 (*ESGF60*) and Estimated Time On Ice Per Game (*ETOIGP),* both did not show any statistically significant increase in draft position. Interestingly, I found that 3rd year draft eligible year CHL Defenseman statistically significantly increase their draft position for all three advanced statistics. Therefore, CHL Defenseman especially for players beginning their hockey career in the CHL or players near the end of eligibility should leverage every opportunity to improve their shooting to evidently increase their Shots Per 60 (*S/60)* advanced statistic. This can be in forms of studying video clips in creating new ways to shoot the puck, dissecting clips of past drafted CHL Defenseman and in conjunction, practicing these clips on the ice. It is important for family advisors, scouts or coaches to relay this information to their clients and utilize this advanced statistic as a measurement for improvement.

Further research needs to be conducted due to the data set I had to work with. This can be done by either collecting more reliable that include all advanced statistics for all observations statistics across more years instead of just the 6. This was one of my biggest limitations to my data as I had to re-work my data set in order to work within the constraints of the Heckman Selection Model. Future researchers in this topic may also want to include various other advanced statistics such as puck possession metrics — Fenwick and Corsi — to better capture the relationship between draft position and advanced statistics for CHL Defenseman. A caveat to these recommendation is that a lot of hockey advanced statistics are hard to obtain across all major junior leagues in the CHL, but they should not be neglected by this fact. With Shots Per 60 (*S/60)* coming out significant, it opens many possibilities to where this field can go.

References:

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

Boost Hockey. (2019). How to Deal With Not Getting Enough Ice Time. Retrieved from: https://www.boosthockey.com/blog/how-to-deal-with-not-getting-enough-ice-time

Brown, M. (2019). What scouts look for — and watch out for — in hockey prospects. Retrieved from: https://www.folio.ca/what-scouts-look-forand-watch-out-forin-hockey-prospects/

Bushway, S., Johnson, B., Slocum, L. (2007). Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *J Quant Criminol,* doi: 10.1007/s10940-007-9024-4

Canucks Army. (2015). Draft Theory: Unveiling the Prospect Cohort Success Model. Retrieved from: https://canucksarmy.com/2015/05/26/draft-analytics-unveiling-the-prospect-cohort-success-model/

Certo, S, T., Busenbark, J, R., Woo, H, S., & Semadeni, M., (2015). Sample Selection Bias And Heckman Models in Strategic Management Research. *Strategic Management Journal, 37(*1), 2639–2657. doi: 10.1002/smj.2475

Couching. (2018). Analytics and the Draft. Retrieved from: https://dobberprospects.com/analytics-and-the-draft/?fbclid=IwAR27qPSX25OrQLPzxcS4fYCubIU0R5cosRg9cvP_Cut347oFnL6akvuZyuE

Desjardins, G. (2004). Hockey analysis and statistics. Retrieved from: http://www.behindthenet.ca/projecting_to_nhl.php

Fischler, S. W., Fischler, S. I., and Eskenazi, G. (2020). Encyclopaedia Britannica: Ice Hockey. Retrieved from: https://www.britannica.com/sports/ice-hockey

Fuchs, J. (2017). NHL's speedy possession game forging a revolution on the blue line. Retrieved from: https://www.si.com/nhl/2015/07/27/puck-possession-game-makes-small-fast-defensemen-most-valuable-in-nhl

Fyffe, I. (2009). Up and coming: Predicting NHL Success. Retrieved from: http://www.hockeyprospectus.com/puck/article.php?articleid=172

Gurra, S, D., Neuberger, J, A. (2010). Economist Incorporated: Sensitivity Analysis in Economic Modeling. Retrieved from: https://ei.com/economists-ink/winter-2010/sensitivity-analysis-in-economic-modeling-by-stuart-d-gurrea-jonathan-a-neuberger/

Hockey Canada (2001). Age Groupings in Minor Will Change for 2002-2003 Season. Retrieved from: https://www.hockeycanada.ca/en-ca/news/2001-nr-057-en

Hurley, W. J., A. Pavlov and W. Andrews. (2012). A Valuation Model for NFL and NHL Draft Positions, *Journal of the Operational Research Society*, *83*(7), 890-98.

Kalchman, L. (2003). Hockey Canada: Making NHL a Very Long Shot. Retrieved from: https://www.hockeycanada.ca/en-ca/news/2003-gn-001-en

Nandakumar, N. (2017). What Does It Mean To Draft Perfectly? An Evaluation Of Draft Strategy In The National Hockey League. *Joseph Warton Research Scholars*. Retrieved from: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1029&context=joseph_wharton_scholars

Puhani, P. (2002). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Survey, 14*(1). 53-68 https://doi-org.ezproxy.library.ubc.ca/10.1111/1467-6419.00104

Schuckers, M. (2016). Draft by numbers: Using data and analytics to improve National Hockey League (NHL) player selection, MIT Sloan Sports Analytics Conference, Boston, 2016. Boston: MIT.

Seppa, T. (2011). Core Age and the Strategic Direction of NHL Teams. In T. Seppa (Ed.), Hockey Prospectus 2011-12 (pp. 404–410).

Sprigings, D. (2016). NHL Draft Probability Tool: Hockey Graphs Retrieved from: https://hockey-graphs.com/2016/06/08/nhl-draft-probability-tool/

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.

The Canadian Hockey League. (n.d.). Kubota Top Prospects. Retrieved from: https://chl.ca/

The Blue Line Hockey I.Q Program, (n.d.). What is Hockey I.Q? Retrieved from: https://www.bluelinehockeyiq.com/hockeyiq

Weissbock, J. (2015). The NHL has a scouting problem. Retrieved from: http://thehockeywriters.com/the-nhl-has-a-scouting-problem/

Weissbock, J. (2015). Draft Analytics: Unveiling the prospect cohort success model. Retrieved from: http://canucksarmy.com/2015/5/26/draft-analyticsunveiling-the-prospect-cohort-success-model