

BACHELORARBEIT

*Der Einfluss von Leistungskennzahlen auf die Löhne von Spielern
in der National Hockey League*

Eine Analyse anhand von OLS-Regressionen mit Python

**Zürcher Hochschule für Angewandte Wissenschaften
School of Management and Law**

Eingereicht von: Fabian Johannes

Matrikelnummer: 16-567-034

Eingereicht bei: Dr. Oliver Bachmann

Abgabedatum: 17. Mai 2019

Management Summary

Die National Hockey League (NHL) gehört zu den vier grossen Sportligen in Nordamerika und setzt jährlich Millionenbeträge um. Die Attraktivität der NHL macht aus den professionellen Eishockeyspielern Sportmillionäre. Jonathan Toews von den Chicago Blackhawks verdiente in der Saison 2017/2018 mit 10.5 Mio. US-Dollar am meisten. Da die Spannweite der Löhne in der NHL allerdings sehr gross ist, stellt sich die Frage, welche Faktoren diese Unterschiede begründen. Hierzu liefern unzählige Statistiken über individuelle Spielerleistungen während einer Saison erste Anhaltspunkte. Da in der NHL eine absolute Lohntransparenz herrscht, sind die Gehälter aller Spieler öffentlich einsehbar.

Die vorliegende Arbeit versucht, mithilfe von statistischen Leistungskennzahlen der Saisons 2016/2017 und 2017/2018 die Löhne von Feldspielern und Torhütern der NHL durch die Modellierung von linearen Regressionen zu erklären. Der Fokus wird dabei bewusst auf quantitativ messbare Statistiken über individuelle Leistungen gelegt. Qualitative, schwer messbare Eigenschaften von Spielern wie Führungsqualitäten oder Disziplin werden nicht berücksichtigt.

Nach der Einbettung der Forschungsfrage in den sportökonomischen Kontext und der Erläuterung der praktischen Relevanz folgt ein Literatur Review über die bisherige Forschung. Erläutert wird der Zusammenhang von Leistungskennzahlen und Löhnen von Sportlern, die Erkenntnisse über Lohndiskriminierung in der NHL sowie die Grundprinzipien zur angewandten Methodik. Danach werden die gesammelten Datensätze über Feldspieler und Torhüter beschrieben, bevor die Anwendung der linearen multiplen Regression erklärt wird. Basierend auf den Erkenntnissen aus dem Literatur Review werden dann die als signifikant vermuteten Leistungskennzahlen definiert. Die Datenanalyse umfasst Histogramme über die ausgewählten Statistiken und lineare Einfachregressionen. Multiple lineare Regressionen geben Aufschluss über die Signifikanz der analysierten Leistungskennzahlen und zeigen auf, welche Variablen einen wie grossen Einfluss auf die Löhne der NHL-Spieler haben. Abschliessend werden bestehende Verträge durch die modellierten Regressionen analysiert, ihre praktischen Anwendungsmöglichkeiten erläutert, sowie die Erkenntnisse kritisch gewürdigt.

Die multiplen Regressionen zur Beantwortung der Forschungsfrage werden nach Spielerpositionen und Spielertypen durchgeführt. Für die Auswertung nach Spielerpositionen

werden Regression für Mittelstürmer, linke Flügelstürmer, rechte Flügelstürmer, Verteidiger und Torhüter modelliert. Die Auswertung nach Spielertypen besteht aus Regressionen für Scorers, Enforcers, Grinders, verteidigende Scorers und Aggressors. Dabei wird analysiert, ob die Einteilung nach Spielertypen bessere Informationen über das Verhalten der Saläre liefert, als eine Kategorisierung nach Positionen. Die Ergebnisse zeigen, dass das Alter, die Position im Entry-Draft, die Anzahl bestrittener Spiele in der Regular Season sowie die Anzahl Skorerpunkte einen signifikanten Einfluss auf das Salär von NHL-Spieler haben. Die Plus/Minus-Bilanz, die Anzahl Strafminuten, die Schussquote und das Gewicht erweisen sich als nicht oder nur teilweise signifikant. Die Gehälter für Torhüter werden durch ein zunehmendes Alter, eine hohe Anzahl absolvierter Partien und eine gute Position im Entry-Draft positiv beeinflusst. Paradoxe Weise wirkt sich eine höhere Anzahl an Spielen ohne Gegentor negativ auf den Lohn aus.

Inhaltsverzeichnis

Tabellenverzeichnis	I
Abbildungsverzeichnis	II
Formelverzeichnis	III
Abkürzungsverzeichnis	IV
1 Einleitung	1
1.1 Ausgangslage	1
1.2 Forschungsfrage	2
1.3 Relevanz	2
1.4 Aufbau der Arbeit	3
2 Literatur Review	4
2.1 Der Einfluss von Leistungskennzahlen auf NHL-Löhne	4
2.2 Lohndiskriminierung in der NHL	7
2.3 Spielertypen	7
2.4 Methodik Review	9
3 Datenbeschreibung	10
3.1 Feldspieler	10
3.2 Torhüter	13
4 Methodik	16
4.1 Theoretische Grundlagen der multiplen linearen Regression	16
4.2 Analyse von Leistungskennzahlen und Spielerlöhnen	18
5 Datenanalyse	19
5.1 Relevante Leistungskennzahlen	20
5.1.1 Identifikation der Regressoren	20
5.1.1.1 Feldspieler	20
5.1.1.2 Torhüter	22
5.1.2 Histogramme & lineare Einfachregressionen der Regressoren	22
5.1.2.1 Histogramme	23
5.1.2.2 Regressoren für Feldspieler & Torhüter	25
5.1.2.3 Regressoren für Feldspieler	26
5.1.2.4 Regressoren für Torhüter	29
5.1.3 Arithmetische Mittelwerte ausgewählter Variablen	30
5.1.4 Häufigkeitsverteilung der Spielerlöhne	33
5.2 Regressionen nach Spielerpositionen	33
5.3 Regressionen nach Spielertypen	34
5.3.1 Kategorisierung Stürmer	35

5.3.2	Kategorisierung Verteidiger.....	35
5.3.3	Regressionen	36
5.4	Interpretation & Diskussion der Resultate.....	37
6	Anwendung.....	42
6.1	Analyse bestehender Verträge	42
6.2	Anwendungsmöglichkeiten in der Praxis & kritische Würdigung.....	44
7	Literaturverzeichnis.....	45
8	Anhang	47

Tabellenverzeichnis

#	Bezeichnung	Seite
Tabelle 1:	Analysierte Variablen für Feldspieler	10
Tabelle 2:	Analysierte Variablen für Torhüter	13
Tabelle 3:	Regressoren für die Lohnanalyse von Feldspielern	21
Tabelle 4:	Regressoren für die Lohnanalyse von Torhütern	22
Tabelle 5:	Durchschnitte & prozentuale Anteile der Variablen nach Position	31
Tabelle 6:	<i>p</i> -Werte der Regressionen nach Positionen	34
Tabelle 7:	Kategorisierung der Angreifer nach Spielertypen	35
Tabelle 8:	Kategorisierung der Verteidiger nach Spielertypen	36
Tabelle 9:	<i>p</i> -Werte der Regressionen nach Spielertypen	37
Tabelle 10:	Übersicht der signifikanten Variablen für alle Regressionen	37
Tabelle 11:	Signifikante Variablen und ihre Einflussstärke auf den Lohn	39
Tabelle 12:	Anwendung der Regressionen (RW, S) auf Patrick Kane (CHI)	42
Tabelle 13:	Anwendung der Regressionen (C, S) auf Connor McDavid (EDM)	43

Abbildungsverzeichnis

#	Bezeichnung	Seite
Abbildung 1:	Histogramme der Regressoren für Feldspieler	23
Abbildung 2:	Histogramme der Regressoren für Torhüter	24
Abbildung 3:	Einfache Linearregressionen LNSAL/AGE für alle Spielerpositionen	25
Abbildung 4:	Einfache Linearregressionen LNSAL/DRAFT für alle Spielerpositionen	25
Abbildung 5:	Einfache Linearregressionen LNSAL/GP für alle Spielerpositionen	26
Abbildung 6:	Einfache Linearregressionen LNSAL/WEIGHT für Feldspieler	26
Abbildung 7:	Einfache Linearregressionen LNSAL/PTS für Feldspieler	27
Abbildung 8:	Einfache Linearregressionen LNSAL/PLUSMINUS für Feldspieler	27
Abbildung 9:	Einfache Linearregressionen LNSAL/PIM für Feldspieler	28
Abbildung 10:	Einfache Linearregressionen LNSAL/SP für Feldspieler	28
Abbildung 11:	Einfache Linearregressionen LNSAL/HES für Feldspieler	29
Abbildung 12:	Einfache Linearregressionen LNSAL/HEIGTH, LNSAL/W, LNSAL/SPE & LNSAL/GSAA für Torhüter	29
Abbildung 13:	Einfache Linearregressionen LNSAL/MIN, LNSAL/SO, LNSAL/PPSP für Torhüter	30
Abbildung 14:	Histogramm SAL	33
Abbildung 15:	Histogramm LNSAL	33

Formelverzeichnis

#	Bezeichnung	Seite
Formel 1:	Modell der multiplen linearen Populationsregression	16
Formel 2:	Modell der multiplen linearen Stichprobenregression	17
Formel 3:	Hypothesentestdesign von Stichprobenregressionskoeffizienten	17
Formel 4:	Angewandte multiple lineare Regression (LNSAL)	19

Abkürzungsverzeichnis

Abkürzung	Beschreibung
A	Erzielte Assists ¹ in der Regular Season ²
AGE	Alter in ganzen Jahren am Ende jeweiligen Saison (SEASON)
ANA	Anaheim Ducks
ARI	Arizona Coyotes
ATOI	Durchschnittliche Einsatzzeit pro Spiel in der Regular Season in Minuten
BES	Geblockte Schüsse bei numerischer Gleichheit ³
BOS	Boston Bruins
BUF	Buffalo Sabres
C	Mittelstürmer (Center)
CAR	Carolina Hurricanes
CBJ	Columbus Blue Jackets
CGY	Calgary Flames
CHI	Chicago Blackhawks
COL	Colorado Avalanche
D	Verteidiger (Defense)
DAL	Dallas Stars
DET	Detroit Red Wings
DRAFT	Runde im Entry-Draft, in welcher der Spieler gezogen wurde
EDM	Edmonton Oilers
ESG	Tore bei numerischer Gleichheit in der Regular Season
EVA	Assists bei numerischer Gleichheit in der Regular Season
FLA	Florida Panthers
FLES	Verlorene Bullys ⁴ bei numerischer Gleichheit in der Regular Season
FWES	Gewonnene Bullys bei numerischer Gleichheit in der Regular Season
FWPES	Gewonnene Bullys in Prozent bei Ausgeglichenheit in der Regular Season
G	Erzielte Tore in der Regular Season
GA	Zugelassene Gegentore in der Regular Season
GAA	Durchschnittlich zugelassene Gegentore pro Spiel in der Regular Season
GK	Torhüter (Goalkeeper)
GM	General Manager

¹ Letzter und zweitletzter Pass der Mitspieler vor einem erzieltem Tor

² Umfasst 82 Spiele pro Mannschaft um sich für das Playoff zu qualifizieren

³ Jedes Team hat gleich viele Spieler auf dem Eis (entweder je fünf, je vier oder je drei)

⁴ Anspiel beim Eishockey, ausgeführt durch die beiden Mittelstürmer der Teams

GP	Absolvierte Spiele in der Regular Season
GSAA	Zugelassene Gegentore über dem Durchschnitt in der Regular Season
GWG	Spielentscheidende Tore in der Regular Season
HEIGHT	Körpergrösse in Zentimeter
HES	Ausgeführte Checks bei numerischer Gleichheit in der Regular Season
ID	Spieler-ID
L	Verlorene Spiele in der Regular Season
LAK	Los Angeles Kings
LNSAL	Natürlicher Logarithmus des Jahreslohns in Mio. US-Dollar (Cap Hit)
LW	Linker Flügel (Left Wing)
MIN	Gespielte Minuten (Karrieretotal) in der Regular Season
MIN	Minnesota Wild
MTL	Montreal Canadiens
NHL	National Hockey League
NJD	New Jersey Devils
NSH	Nashville Predators
NYI	New York Islanders
NYR	New York Rangers
OLS	Ordinary Least Squares (Regressionsmethode der kleinsten Quadrate)
OTL	Niederlagen nach Verlängerung oder Penaltyschiessen in der Regular Season
OTT	Ottawa Senators
OTW	Siege in der Verlängerung (Overtime Wins)
PA	Erzielte Assists in den Playoffs ⁵
PATOI	Durchschnittliche Einsatzzeit pro Spiel in den Playoffs in Minuten
PBES	Geblockte Schüsse bei numerischer Gleichheit in den Playoffs
PESG	Tore bei numerischer Gleichheit in den Playoffs
PEVA	Assists bei numerischer Gleichheit in den Playoffs
PFLES	Verlorene Bullys bei numerischer Gleichheit in den Playoffs
PFWES	Gewonnene Bullys bei numerischer Gleichheit in den Playoffs
PFWPES	Gewonnene Bullys in Prozent bei numerischer Gleichheit in den Playoffs
PG	Erzielte Tore in den Playoffs
PGA	Zugelassene Gegentore in den Playoffs
PGAA	Durchschnittlich zugelassene Gegentore pro Spiel in den Playoffs
PGP	Absolvierte Spiele in den Playoffs

⁵ 16 von 31 Teams qualifizieren sich und spielen um den Gewinn der Meisterschaft (Stanley Cup)

PGSAA	Zugelassene Gegentore über dem Durchschnitt in den Playoffs
PGWG	Spielentscheidende Tore in den Playoffs
PHES	Ausgeführte Checks bei numerischer Gleichheit in den Playoffs
PHI	Philadelphia Flyers
PIM	Anzahl Strafminuten in der Regular Season
PIMPGP	Anzahl Strafminuten pro Spiel in der Regular Season
PIT	Pittsburgh Penguins
PL	Verlorene Spiele in den Playoffs
PLAYER	Vorname und Name des Spielers
PLUSMINUS	Plus/Minus-Statistik in der Regular Season
PLUSMINUSPGP	Plus/Minus-Statistik pro Spiel in der Regular Season
PMIN	Gespielte Minuten (Karrieretotal) in den Playoffs
POS	Spielerposition
POTL	Niederlagen nach Verlängerung oder Penaltyschiessen in den Playoffs
PPA	Powerplay ⁶ -Assists in der Regular Season
PPG	Powerplay-Tore in der Regular Season
PPGA	Zugelassene Gegentore bei gegnerischem Powerplay in der Regular Season
PPIM	Anzahl Strafminuten in den Playoffs
PPLUSMINUS	Plus/Minus-Statistik in den Playoffs
PPPA	Powerplay-Assists in den Playoffs
PPPG	Powerplay-Tore in den Playoffs
PPSA	Erhaltene Torschüsse bei gegnerischem Powerplay in der Regular Season
PPSP	Fangquote in Prozent bei gegnerischem Powerplay in der Regular Season
PPTS	Erzielte Punkte (Tore und Assists) in den Playoffs
PRF	Populationsregressionsfunktion
PSA	Erhaltene Torschüsse in den Playoffs
PSHA	Unterzahl-Assists in den Playoffs
PSHG	Unterzahl-Tore in den Playoffs
PSO	Absolvierte Spiele ohne Gegentor (Shutouts) in den Playoffs
PSOG	Total der Torschüsse in den Playoffs
PSP	Erfolgreiche Torschüsse in Prozent in den Playoffs
PSPE	Fangquote in Prozent in den Playoffs
PSV	Gehaltene Torschüsse in den Playoffs
PTOI	Einsatzzeit in Minuten in den Playoffs

⁶ Überzahlsituation aufgrund einer Strafe für das gegnerische Team

PTS	Erzielte Punkte (Tore und Assists) in der Regular Season
PTSPGP	Anzahl Skorerpunkte pro Spiel in der Regular Season
PTSPGP	Anzahl Skorerpunkte pro Spiel in den Playoffs
PW	Gewonnen Spiele in den Playoffs
RSTOI	Einsatzzeit in Minuten in der Regular Season
RW	Rechter Flügel (Right Wing)
SA	Erhaltene Torschüsse in der Regular Season
SAL	Jahreslohn in Mio. US-Dollar (Cap Hit)
SEASON	Saison, aus welcher die Leistungskennzahlen stammen
SHA	Unterzahl-Assists in der Regular Season
SHG	Unterzahl-Tore in der Regular Season
SHGA	Zugelassene Gegentore bei gegnerischem Boxplay ⁷ in der Regular Season
SHSA	Erhaltene Torschüsse bei gegnerischem Boxplay in der Regular Season
SHSP	Fangquote in Prozent bei gegnerischem Boxplay in der Regular Season
SJS	San Jose Sharks
SO	Absolvierte Spiele ohne Gegentor (Shutouts) in der Regular Season
SOG	Torschüsse in der Regular Season
SOP	Schusshand des Spielers, entweder links oder rechts
SP	Erfolgreiche Torschüsse in Prozent in der Regular Season
SPE	Fangquote in Prozent in der Regular Season
SRF	Stichprobenregressionsfunktion
STAR	Teilnahme an den Allstar-Games der jeweiligen Saison (SEASON)
STL	St. Louis Blues
SV	Gehaltene Torschüsse in der Regular Season
TBL	Tampa Bay Lightning
TEAM	NHL-Team, das den Spieler unter Vertrag hat
TOR	Toronto Maple Leafs
VAN	Vancouver Canucks
VGK	Las Vegas Golden Knights
W	Gewonnene Spiele in der Regular Season
WEIGHT	Körpergewicht in Kilogramm
WHA	World Hockey Association
WPG	Winnipeg Jets
WSH	Washington Capitals

⁷ Unterzahlsituation aufgrund einer Strafe für das eigene Team

1 Einleitung

Eishockey ist für viele Leute eine faszinierende, schnelle und emotionsgeladene Sportart. Eishockey wurde – wie viele andere Sportarten auch – in den letzten Jahrzehnten extrem kommerzialisiert.

Der Artikel von Ozeanian und Badenhausen (2018) über finanzielle Kennzahlen der nordamerikanischen National Hockey League (NHL) für die Saison 2017/2018 belegt diese Entwicklung eindrücklich. Der durchschnittliche Umsatz der 31 Teams betrug 157 Millionen US-Dollar, der Betriebsgewinn lag im Mittel bei 25 Millionen US-Dollar. Letztere Kennzahl hat sich seit der Spielzeit 2011/2012 verdreifacht. Der Gesamtwert der Mannschaften wurde auf 19.5 Milliarden US-Dollar geschätzt. Das entspricht einem durchschnittlichen Wertzuwachs von sieben Prozent gegenüber der Vorsaison.

Es sei an dieser Stelle erwähnt, dass ein Grossteil der Wertentwicklungen Schätzungen sind, da ausser den New York Rangers, den Toronto Maple Leafs und den Montreal Canadiens kein NHL-Team im Besitz einer Aktiengesellschaft ist. Daher ist ein Grossteil der Teams von der Pflicht, einen Geschäftsbericht zu veröffentlichen, entbunden (InvestSnips, 2017).

1.1 Ausgangslage

Nutzniesser einer gewinnbringenden Eishockeyliga in Nordamerika sind unter anderem die Spieler der Teams, die zu den am besten bezahlten Sportlern der Welt zählen. Die Saläre der Akteure variieren dennoch enorm. Gemäss Vollman (2018) verdiente der 29-jährige Paul Carey, Mittelstürmer der Ottawa Senators, in der Saison 2017/2018 insgesamt 650'000 US-Dollar. Jonathan Toews von den Chicago Blackhawks, ebenfalls 29 Jahre alt und Mittelstürmer, erhielt in derselben Spielzeit eine Vergütung von 13.8 Mio. US-Dollar (Vollman, 2018). Da beide Spieler gleich alt sind und dieselbe Position innehaben, erscheint es logisch, dass aufgrund des höheren Lohnes die Fähigkeiten von Jonathan Toews höher einzustufen sind, als jene von Paul Carey. Der angenommene Leistungsunterschied zwischen diesen beiden Spielern kann mit einem Vergleich von Leistungskennzahlen (absolvierte Spiele, erzielte Tore, erzielte Assists, Anzahl Strafminuten etc.) begründet werden. Welche Einflussfaktoren genau erklären aber eine Lohnspannweite von 13.15 Mio. US-Dollar in der besten Eishockeyliga der Welt?

1.2 Forschungsfrage

Einzelleistungen von Spielern lassen sich im Eishockey mithilfe von quantitativen Variablen beschreiben, wovon einige signifikanter für einen Teamerfolg sind und andere weniger. Hinzu kommen viele qualitative, nicht messbare Einflüsse, die die Leistung eines Eishockeyspielers prägen. Die psychische und physische Verfassung, Führungsqualitäten oder die Arbeitsmentalität um nur einige zu nennen.

Diese Arbeit soll untersuchen, wie die quantitativ messbaren Leistungskennzahlen die Löhne von Spielern in der National Hockey League beeinflussen. Die qualitativen Faktoren und mögliche Synergieeffekte, die sich aus der Zusammenstellung der Mannschaft ergeben, werden bewusst ausgeklammert. Der Grund dafür liegt in der beinahe unmöglichen Quantifizierung dieser Grössen.

1.3 Relevanz

Gemäss dem allgemeinen sportökonomischen Sachverstand hat in der kommerzialisierten Welt des Sports das Team mit den grössten finanziellen Möglichkeiten die grössten Chancen auf den Gewinn einer Meisterschaft. Dieses Phänomen ist eindrücklich in den grössten europäischen Fussballligen zu beobachten.

Damit das gleiche Schicksal der NHL nicht wiederfährt, wurde ab der Saison 2004/2005 ein Salary Cap eingeführt. Der Salary Cap besteht aus einer Gehaltsobergrenze und einer Gehaltsuntergrenze der totalen Lohnausgaben eines Teams in einer Saison (Hofer, 2017). Hofer (2017) ergänzt, dass der Salary Cap für alle Mannschaften verbindlich ist und er jede Saison neu festgelegt wird. So soll sichergestellt werden, dass trotz unterschiedlichen finanziellen Möglichkeiten die Konkurrenzfähigkeit erhalten bleibt.

Der Salary Cap erfordert von den General Managers (GM) der NHL-Teams, mit dem durch die Gehaltsgrenzen definierten Budget eine konkurrenzfähige Mannschaft zusammenzustellen. Wenn die Löhne von NHL-Spielern zu einem grossen Teil durch quantitative Leistungszahlen erklärt werden können, sind GM bei der Zusammenstellung ihres Kaders in der Lage, auf diese Zusammenhänge zurückzugreifen. Dies würde eine effizientere Nutzung der Lohnsumme zur Folge haben. Eine Analyse des Zusammenhangs zwischen Leistungskennzahlen und Löhnen kann helfen, überbezahlte Verträge neu auszuhandeln oder unterbezahlte Spieler eines Konkurrenten mit einem besseren Angebot zu

verpflichten. So kann ein GM mithilfe der Statistik sein Team verstärken, was die Erfolgsaussichten auf den Gewinn des Stanley Cups⁸ erhöht.

1.4 Aufbau der Arbeit

Die Einleitung erörterte den möglichen Zusammenhang zwischen Leistungskennzahlen von NHL-Spielern und ihren Löhnen. Im Allgemeinen wurde auf die praktische Relevanz für die General Managers eingegangen, im Speziellen auf die Implikationen der Lohnforschung in Bezug auf den Salary Cap.

In Kapitel 2 folgt ein Literatur Review. Abschnitt 2.1 zeigt auf, von welchen Komponenten der Lohn im Generellen abhängt und welche Erkenntnisse die bisherige Forschung bezüglich Lohndeterminanten von NHL-Spielern erlangte. Darüber hinaus wird das Thema der Lohndiskrimination angeschnitten (2.2). Eine Arbeit über verschiedene Spielertypen könnte eine interessante Basis für alternative Regressionsmodelle darstellen, weshalb ihre Ergebnisse in Kapitel 2.3 erläutert werden. Der Methodik Review (2.4) fasst Erkenntnisse über das Vorgehen der bisherigen Forschungen im Bereich der Lohnanalyse zusammen. Sie dienen als Grundlage, um das Vorgehen dieser Arbeit zu strukturieren.

Kapitel 3 beschreibt die Charakteristika der erhobenen Datensätze und erklärt die erfassten Leistungskennzahlen für Feldspieler (3.1) und Torhüter (3.2).

Kapitel 4 gibt eine Übersicht über die Theorie der multiplen linearen Regression (4.1) und ihre Anwendung auf die Forschungsfrage (4.2).

Das fünfte Kapitel analysiert die im 3. Abschnitt beschriebenen Datensätze über Feldspieler und Torhüter der Saison 2016/2017 und 2017/2018. Zuerst werden mögliche signifikanten Variablen für Feldspieler und Torhüter identifiziert (5.1.1). Danach untersucht Kapitel 5.1.2 das Verhalten der Regressoren mit Histogrammen und linearen Einfachregressionen zum logarithmierten Salär der Spieler. Weiter geben arithmetische Mittelwerte über ausgewählten Leistungskennzahlen (5.1.3) eine detailliertere Übersicht über die Datensätze. In Kapitel 5.1.4 wird die Verteilung der Spielerlöhne und die daraus entstehenden Implikationen für die Regressionskonstruktionen diskutiert. Anschliessend folgen die Regressionen nach Spielerpositionen (5.2) und Spielertypen (5.3). Letzteres Kapitel zeigt

⁸ Meisterschaftstrophäe der NHL

zuerst die Kategorisierung von Stürmer (5.3.1) und Verteidiger (5.3.2), bevor pro Kategorie eine Regression modelliert wird (5.3.3). Abschliessend werden in Kapitel 5.4 die Regressionsresultate interpretiert und diskutiert.

Schliesslich werden zwei bestehende NHL-Verträge mithilfe der konstruierten Regressionen analysiert (6.1). Kapitel 6.2 zeigt Anwendungsmöglichkeiten der Forschungsergebnisse auf und würdigt diese kritisch.

2 Literatur Review

Der Literatur Review zeigt den aktuellen Forschungsstand zum Zusammenhang von Leistungskennzahlen und Löhnen in der NHL auf und schneidet das Thema der Lohndiskriminierung an. Die Forschung über verschiedene Spielertypen könnte die Basis für eine alternative Regressionskonstruktion sein, weshalb ihre Ergebnisse erläutert werden. Der Methodik Review fasst das Vorgehen der bisherigen Forschungen zusammen.

2.1 Der Einfluss von Leistungskennzahlen auf NHL-Löhne

Jacob Mincer (1958, S. S.281) legte in seiner Forschung über Investitionen in das Humankapital dar, dass das Einkommen einer Person von ihren Fähigkeiten und ihrer Erfahrung abhängt. Je höher die Fähigkeiten und die Erfahrung, desto grösser die Produktivitätsvorteile für die Firma und damit das Salär für die Person. Als Fähigkeit wird dabei die Schulbildung als zentral betrachtet. Ab einem gewissen Alter ist das Salär abnehmend, da durch das zunehmende Alter die Produktivität abnimmt. Gemäss Vincent und Eastman (2009b, S. 256) hat die Mincer-Einkommensgleichung eine geringere Aussagekraft für schlechtbezahlte Spieler der NHL als für Athleten mit höher dotierten Verträgen. Laut Wiseman und Chatterjee (2010, S. 1) sind die Löhne von Athleten generell von ihren bisherigen Leistungen und ihrem Potential abhängig.

Eastman (1981, S. 100) hat die Mincer-Einkommensgleichung erstmals auf die Löhne von 44 NHL-Spielern der Saison 1976/1977 angewandt. Die in der Mincer-Einkommensgleichung beschriebenen Einflussfaktoren der Fähigkeiten (Skorerpunkte) und der Erfahrung hatten dabei einen signifikanten positiven Effekt auf den Lohn eines Eishockeyspielers. Weiter wirkte sich ein zunehmendes Alter uneingeschränkt positiv auf den Lohn aus, nämlich jährlich um 20'650 US-Dollar. In Kontrast zur Mincer-Einkommensgleichung schlug sich eine längere Schulbildung negativ auf den Lohn von NHL-Spielern nieder.

Die Abnahme betrug 16'000 US-Dollar pro zusätzliches Schuljahr. Andere Einflussfaktoren erwiesen sich als insignifikant.

Longley (1995, S. 416) verglich die Skorerwerte von 250 Offensivspielern der NHL der Saison 1989/1990 mit ihren Löhnen. Die Regression beinhaltete zudem Daten über die Anzahl gespielter Partien, die Gesamteinnahmen der jeweiligen Mannschaften sowie die Herkunft der Spieler. Dabei erwiesen sich Skorerpunkte, Teameinnahmen und gespielte Partien als positive Lohneinflüsse. Ein negativer Faktor sei eine frankokanadische Herkunft, die einen Abschlag von 31-41 Prozent ausmache. Detailliertere Ausführungen zur Lohndiskriminierung in der NHL folgen in Kapitel 2.2. Kahane (2001, S. 632) zeigte ebenfalls auf, dass höhere Einnahmen einer Franchise zu höheren Spielerlöhnen führten. Lohnunterschiede aufgrund von finanziellen Möglichkeiten der Mannschaften wurden durch die Einführung des Salary Caps nach der Lockout-Saison⁹ 2004/2005 eliminiert. Der Salary Cap strukturiert die Lohnbudgets, die von Spielern und Management genehmigt werden (Wiseman & Chatterjee, 2010, S. 1).

In zahlreichen bisherigen Forschungen wurde die Plus/Minus-Statistik eines Spielers als signifikantes Charakteristikum für die Erklärung seines Salärs aufgeführt. Vincent und Eastman (2009b, S. 261) definieren die Plus/Minus-Statistik wie folgt:

“The plus-minus statistic is a measure of a player’s two-way play: it is the difference between the number of times the player is on the ice when his team scores a goal and the number of times he is on the ice when an opposing team scores a goal. In both cases, the teams are at even strength”.

Gramacy, Jensen und Taddy (2012, S. 2) bezeichnen die Plus/Minus-Statistik aufgrund ihrer Einfachheit als wichtigste Grösse, um die Leistung eines Spielers zu beurteilen. Allerdings führen sie aus, dass Plus/Minus-Statistiken erheblich von Mitspielern und Gegnern abhängig seien, weshalb diese Werte kritisch betrachtet werden sollten. Weiter argumentieren sie aus, dass die Kennzahl bei denjenigen Spielern nicht aussagekräftig sei, die wenig Einsatzzeit (TOI) erhalten. Kritisch betrachtet auch Walsh (1992, S. 445) die Plus/Minus-Statistik, insbesondere befindet er, dass sie keine Aussage über die Defensivfähigkeiten eines Spielers aussagen würde. Auch Lavoie, Grenier und Coulombe (1992, S. 462) bemängeln die Aussagekraft der Plus/Minus-Statistik. Sie argumentieren, dass

⁹ Verkürzte oder abgesagte Saison aufgrund rechtlicher Streitereien zwischen Liga und Spielergewerkschaften

die Anzahl erhaltenen Gegentore im Boxplay eine bessere Aussagekraft über die Verteidigungsqualitäten eines Spielers aufweise. Begründet wird die Aussage mit der Annahme, dass ein Trainer seine besten Defensivspieler auf das Eis beorderte, wenn seine Mannschaft mit einem Spieler weniger agieren müsse. Übersteht ein Spieler viele Einsätze im Boxplay ohne ein Gegentor zu erhalten, seien seine Verteidigungsqualitäten als hoch einzustufen.

Jones, Nadeau und Walsh (1997, S. 204) untersuchten, ob sich die Anzahl Strafminuten positiv oder negativ auf das Salär eines NHL-Spielers auswirken. Sie sagen, dass prügelnde Spieler zahlende Gäste in die Stadien locken würden und sie daher für ihr Talent bezahlt werden sollten, wie Skorer für das Erzielen von Toren bezahlt würden. Der vermutete Kausalzusammenhang von Aggressivität und höherem Lohn konnte indes nicht bewiesen werden. Generell scheint der Einfluss von Strafminuten ambig zu sein. Viele Strafminuten wirken sich negativ auf die Leistung von gut bezahlten Leistungsträger aus, jedoch positiv auf die Leistung von weniger gut bezahlten Spielern, deren Hauptaufgabe es sei, den Gegner einzuschüchtern (Vincent & Eastman, 2009b, S. 257).

Nebst Skorerwerten, Plus/Minus-Statistiken und Strafminuten wurden weitere Leistungskennzahlen auf ihre Signifikanz hinsichtlich des Salärs untersucht. Walsh (1992, S. 444) argumentierte, dass die Körpergrösse zwar nicht positiv mit einem höheren Gehalt korreliere, sie aber ein kritischer Faktor darstelle, um überhaupt eine Karriere in der NHL einschlagen zu können.

Jones und Walsh (1988, S. 602) untersuchten den Einfluss einer Konkurrenzliga auf die Gehälter von NHL-Spielern. Sie entdeckten, dass in den Jahren 1972-1979 die Löhne von NHL-Spielern markant angestiegen sind, und stellten einen Kausalzusammenhang mit der Existenz der Konkurrenzliga WHA (World Hockey Association) dar, die in diesem Zeitraum das Monopol der NHL aufzulösen versuchte. In der heutigen Zeit scheint der Einflussfaktor anderer Ligen auf die Spielerlöhne irrelevant, da das Monopol der NHL als beste Eishockeyliga der Welt gemäss der allgemeingültigen Meinung als unbestritten gilt. Die Monopolstellung der National Hockey League kann weiter durch das Draft-System begründet werden. Der Mechanismus räumt den NHL-Teams das Recht ein, weltweit die besten Talente vertraglich an sich zu binden, sollten sie jemals in der NHL spielen (Eishockey.com, 2019). Dieses Recht impliziert, dass die NHL als beste Liga der Welt angesehen wird. Laut Vincent und Eastman (2009b, S. 275) würde eine höhere Position im Entry Draft einen signifikanten positiven Effekt auf das Salär haben.

Idson und Kahane (2000, S. 356) argumentierten, dass der Lohn eines Eishockeyspielers nicht von seinen Fähigkeiten alleine abhängig sei, sondern in Perspektive zum Talent seiner Mitspieler betrachtet werden müsse. Besitze ein Spieler ein ausgezeichnetes Defensivverhalten, sei sein Team nur bereit ein höheres Gehalt zu bezahlen, falls die Mannschaft nicht bereits über genügend Spieler mit guten Verteidigungsqualitäten verfüge.

Zusammenfassend ist sich die bisherige Forschung einig, dass Skorerpunkte, die Plus/Minus-Statistik und Strafminuten einen signifikanten Einfluss auf die Höhe des Gehalts hätten. Dies wurde durch die Analysen von Löhnen im Laufe der Zeit bewiesen (Eastman, 1981; Longley, 1995; Vincent & Eastman, 2009b). Konträr dazu ist durch die Einführung des Salary Caps nach der Saison 2004/2005 der einst signifikante Einfluss der Franchise-Einnahmen auf die Spielergehälter (Kahane, 2001) nicht mehr existent.

2.2 Lohndiskriminierung in der NHL

Nach Mincer (1958, S. 302) implizieren gleiche Fähigkeiten und gleiche Chancen nicht den gleichen Lohn. Wie bereits erwähnt, umfassen die Fähigkeiten eines Eishockeyspielers weit mehr als nur quantitative Leistungskennzahlen. Sie sind somit beinahe unmöglich vergleichbar. Dennoch ist die Lohndiskriminierung in der NHL ein weit erforschtes Phänomen. Um 1970 wurde eine Lohndiskriminierung der Verteidiger festgestellt, die aber bis 1990 wieder abgenommen hat (Lavoie & Grenier, 1992). Die Lohndiskriminierung von frankokanadischen Eishockeyspielern in der NHL wurde wiederholt untersucht. Longley (1995, S. 420) eruierte eine Lohnnivellierung von 33 Prozent für Frankokanadier, die ausserhalb des frankophonen Teils der NHL spielten. Gegenüberstellend verwarfen Jones, Nadeau und Walsh (1999, S. 605) die Hypothese der Lohndiskriminierung von frankophonen Kanadiern. Da offizielle Daten über Spielerlöhne vor 1990 nur für zwei Saisons öffentlich zugänglich sind, können keine fundierten Aussagen über den Trend von möglichen Lohndiskriminierung während jener Zeit gemacht werden (Leadley & Zygmunt, 2006, S. 57).

2.3 Spielertypen

Die Regressionen zur Modellierung von Löhnen und ihren Abhängigkeiten sollten nur Beobachtungen einer in sich homogenen Gruppe enthalten. Lohnvergleiche heterogener Spielertypen machen keinen Sinn und reduzieren das Bestimmtheitsmass der Regression (vgl. 4.1). Es scheint naheliegend, dass man Eishockeyspieler nach ihrer Position (An-

griff, Verteidigung, Torhüter) in homogene Gruppen einteilen kann. Anders haben Vincent und Eastman (2009a) Feldspieler kategorisiert. Sie klassifizierten Stürmer als Scorers, Grinders oder Enforcers. Verteidiger wurden entweder als Scorers oder als Aggressors beschrieben. Die Kategorien wurden anhand von Leistungskennzahlen mehrerer Regular Seasons, bis einschliesslich der Saison 2002/2003, gebildet.

Die Spielerkategorien für Angreifer werden von Vincent und Eastman (2009a, S. 1) wie folgt definiert: Scorers werden auch als «Power Forwards» bezeichnet. Sie lassen sich durch Skorerwerte identifizieren, die über dem Durchschnitt für die jeweilige Position liegen. Ein Grinder ist ein Stürmer, dessen Hauptaufgabe darin besteht, die gegnerischen Angriffe durch ein physisch intensives Spiel zu stoppen und Gegenspieler zu zermürben, also die gegnerischen Scorers aus dem Spiel zu nehmen. Enforcers sind mit Grinders zu vergleichen, haben aber zusätzlich die wichtige Aufgabe, Schlüsselspieler des eigenen Teams zu beschützen, indem sie sich gegen die Grinders des Gegners opponieren. Früher wurden Grinders auch als Goons bezeichnet. Goons lassen sich durch eine überdurchschnittliche Anzahl Strafminuten und anhand von unterdurchschnittlichen Skorerwerten identifizieren (Vincent & Eastman, 2009a, S. 2).

Die Kategorisierung der Angreifer (Vincent & Eastman, 2009a, S. 9) ergab, dass Scorers im Durchschnitt 0.77 Skorerpunkte pro Spiel (PTSPGP) aufweisen und eine Plus/Minus-Statistik von +0.07 pro Spiel haben. 31.3 Prozent der Spieler wurden als Scorers identifiziert. Grinders erzielen im Mittel 0.35 Punkte pro Spiel und haben eine Plus/Minus-Bilanz von -0.04 pro Spiel. Mit 56.9 Prozent bilden sie die grösste Gruppe der Angreifer. Das physisch intensive Spiel der Enforcers schlägt sich in der schlechtesten Puntkeausbeute nieder, die bei durchschnittlich 0.19 Punkten pro Spiel liegt. Die Plus/Minus-Bilanz pro Spiel beträgt im Mittel -0.05. Die Anzahl Strafminuten pro Spiel liegt mit 2.46 Minuten deutlich über dem Wert für Scorers und Grinders. 11.8 Prozent der Angreifer wurden als Enforcers klassifiziert.

Verteidiger wurden von Vincent und Eastman (2009a, S. 16) in zwei unterschiedliche Gruppen eingeteilt. Spielerisch bessere Verteidiger wurden als Scorers identifiziert. Sie identifizieren sich durch bessere Skorerwerte (PTS) und weniger Strafminuten (PIM). Als Aggressors wurden Verteidiger beschrieben, die für die physische Komponente des Spiels zuständig sind. Aggressors sind durchschnittlich schwerer als Scorers. Gemessen an der Variable PTSPGP wurden 55 Prozent der Verteidiger als Scorers klassifiziert. Die restlichen 45 Prozent wurden der Kategorie Aggressors zugeteilt. Die durchschnittliche

Anzahl Skorerpunkte pro Spiel unterscheidet Scorers (0.35 PTSPGP) deutlich von Aggressors (0.15 PTSPGP). Die Plus/Minus-Bilanz pro Spiel der Scorers ist positiv, jene der Aggressors negativ. Zudem sammelten Aggressors deutlich mehr Strafminuten pro Spiel und wiesen im Durchschnitt ein höheres Körpergewicht auf, als Scorers.

2.4 Methodik Review

Die bisherige Forschung über NHL-Löhne in Abhängigkeit von Leistungskennzahlen basiert grösstenteils auf Regressionen nach OLS (Ordinary Least Squares, Methode der kleinsten Quadrate) (Leadley & Zygmunt, 2006, S. 74). Konträr dazu argumentierten Vincent und Eastman (2009b, S. 275), dass eine Regression nach OLS für die Gehaltserklärung von NHL-Spielern nicht ausreiche. Sie erforschten die Zusammenhänge zwischen Leistungskennzahlen und Löhnen mittels Quantil-Regressionen und OLS. Ein Vergleich der beiden Regressionen ergab unter anderem unterschiedliche Signifikanzergebnisse der Plus/Minus-Statistik bei Verteidigern (Vincent & Eastman, 2009b, S. 272). Vincent und Eastman (2009b) inkludierten in ihrer Analyse nur Spieler, die mindestens zehn Spiele in der Regular Season bestritten haben.

Wiseman und Chatterjee (2010, S. 1) argumentieren, dass eine OLS-Regression keine zuverlässigen Resultate liefere, falls die zugrunde liegenden Daten nicht normalverteilt seien. Die von ihnen analysierten Gehältern von Sportlern der Major League Baseball waren nicht normalverteilt. Eine OLS-Regression überschätze im Falle von nicht normalverteilten Daten die Löhne von schlechten Spielern und unterschätze jene von sehr guten Spielern (Wiseman & Chatterjee, 2010, S. 1). Auch Mincer (1958, S. 283) bestätigt, dass Einkommen von Personen keiner Normalverteilung folgen würden, die logarithmierten Löhne allerdings schon. Eine Auswertung von logarithmierten Gehältern wurde unter anderem von Mincer (1958), Eastman (1981), Idson und Kahane (2000) und Vincent und Eastman (2009b) durchgeführt.

Die bisherigen Forschungen zeigen, dass die besten Regressionsmodelle über Löhne von Sportlern relativ einfach aufgebaut sind und Bestimmtheitsmasse von bis zu 80 Prozent erreichen.

3 Datenbeschreibung

Die erhobenen Statistiken stammen aus der Regular Seasons 2016/2017 und 2017/2018 sowie aus den Playoffs 2016/2017 und 2017/2018. Es wurden Daten über Spieler erhoben, die mindestens eine Partie in der jeweiligen Regular Season bestritten haben.

3.1 Feldspieler

Die zu analysierenden Leistungskennzahlen über Feldspieler stammen von Hockey Reference (2018) und Vollman und Cane (2018). Die Spielergehälter wurden von Vollman und Cane (2018) aufbereitet. Kleine Ergänzungen zu fehlenden Löhnen wurden durch Spotrac (2018) vorgenommen. Der Datensatz umfasst 1853 Stichproben, wobei Spieler, die unter der Saison zu einem anderen Team transferiert wurden, einen Eintrag pro Team aufweisen. Um die Korrektheit der Daten zu verifizieren wurden zehn Beobachtungen des Datensets mit den offiziellen Statistiken der NHL verglichen.

In der nachfolgenden Tabelle werden sämtliche Leistungskennzahlen beschrieben, die laut bisheriger Forschung einen Einfluss auf das Salär von Stürmer und Verteidiger haben könnten. Dabei ist zwischen allgemeinen Daten im ersten Abschnitt der Tabelle, Leistungskennzahlen aus der Regular Season im zweiten Teil und Statistiken aus den Playoffs im unteren Drittel zu unterscheiden.

Tabelle 1: Analytierte Variablen für Feldspieler

Variable	Bezeichnung
SAL	Jahreslohn in Mio. US-Dollar (Cap Hit)
LNSAL	Natürlicher Logarithmus des Jahreslohns in Mio. US-Dollar (Cap Hit)
ID	Spieler-ID
PLAYER	Vorname und Name
SEASON	Saison, aus welcher die analysierten Leistungskennzahlen stammen
AGE	Alter in ganzen Jahren am Ende jeweiligen Saison
HEIGHT	Körpergrösse in Zentimeter
WEIGHT	Körpergewicht in Kilogramm
STAR	Boolesche Variable; Wahr, falls Teilnahme an den All-Star Games
DRAFT	Runde, in welcher der Spieler gedraftet wurde
TEAM	NHL-Team, das den Spieler unter Vertrag hat
SOP	Schusshand des Spielers, entweder links oder rechts
POS	Spielerposition
GP	Absolvierte Spiele in der Regular Season

G	Erzielte Tore in der Regular Season
A	Erzielte Assists in der Regular Season
PTS	Erzielte Punkte (Tore und Assists) in der Regular Season
PTSPGP	Erzielte Punkte pro absolviertem Spiel in der Regular Season
PLUSMINUS	Plus/Minus-Statistik in der Regular Season
PIM	Anzahl Strafminuten in der Regular Season
ESG	Tore bei numerischer Gleichheit in der Regular Season
PPG	Powerplay-Tore in der Regular Season
SHG	Untersahl-Tore in der Regular Season
GWG	Spielentscheidende Tore in der Regular Season
EVA	Assists bei numerischer Gleichheit in der Regular Season
PPA	Powerplay-Assists in der Regular Season
SHA	Untersahl-Assists in der Regular Season
SOG	Torschüsse in der Regular Season
SP	Erfolgreiche Torschüsse in Prozent der Anzahl Schüsse in der Regular Season
TOI	Einsatzzeit in Minuten in der Regular Season
ATOI	Durchschnittliche Einsatzzeit pro Spiel in der Regular Season in Minuten
BES	Geblockte Schüsse bei numerischer Gleichheit
HES	Ausgeführte Checks bei numerischer Gleichheit in der Regular Season
FWES	Gewonnene Bullys bei numerischer Gleichheit in der Regular Season
FLES	Verlorene Bullys bei numerischer Gleichheit in der Regular Season
FWPES	Gewonnene Bullys in Prozent bei numerischer Gleichheit in der Regular Season
<hr/>	
PGP	Absolvierte Spiele in den Playoffs
PG	Erzielte Tore in den Playoffs
PA	Erzielte Assists in den Playoffs
PPTS	Erzielte Punkte (Tore und Assists) in den Playoffs
PTSPPGP	Erzielte Punkte pro absolviertem Spiel in den Playoffs
PPLUSMINUS	Plus/Minus-Statistik in den Playoffs
PPIM	Anzahl Strafminuten in den Playoffs
PESG	Tore bei numerischer Gleichheit in den Playoffs
PPPG	Powerplay-Tore in den Playoffs
PSHG	Untersahl-Tore in den Playoffs
PGWG	Spielentscheidende Tore in den Playoffs
PEVA	Assists bei numerischer Gleichheit in den Playoffs
PPPA	Powerplay-Assists in den Playoffs
PSHA	Untersahl-Assists in den Playoffs
PSOG	Total der Torschüsse in den Playoffs

PSP	Erfolgreiche Torschüsse in Prozent in den Playoffs
PTOI	Einsatzzeit in Minuten in den Playoffs
PATOI	Durchschnittliche Einsatzzeit pro Spiel in den Playoffs in Minuten
PBES	Geblockte Schüsse bei numerischer Gleichheit
PHES	Ausgeführte Checks bei numerischer Gleichheit in den Playoffs
PFWES	Gewonnene Bullys bei numerischer Gleichheit in den Playoffs
PFLES	Verlorene Bullys bei numerischer Gleichheit in den Playoffs
PFWPES	Gewonnene Bullys in Prozent bei Ausgeglichenheit in den Playoffs

Die Gehälter (SAL) beinhalten die für den Cap Hit relevanten Löhne, also ohne allfällige Sonderzahlungen wie beispielsweise Signing Fees¹⁰. Um das Problem der nicht normalverteilten Löhne und die damit verbundenen Schwierigkeiten für die Regressionskonstruktion (vgl. 5.1.4) zu lösen, beinhaltet der Datensatz nebst den Löhnen (SAL) die logarithmierten Saläre (LNSAL). Die Variablen ID und PLAYER dienen der eindeutigen Identifikation eines Spielers und seinen Statistiken. Die möglichen Regressoren HEIGHT und WEIGHT wirken sich erwartungshalber positiv auf das Salär aus. Dies könnte durch eine erhöhte physische Durchsetzungskraft im Spiel begründet werden. Die Variable STAR sagt aus, ob ein Spieler für die All-Star Games in der jeweiligen Saison (SEASON) aufgeboden wurde. Für die All-Star Games 2018 wurden 37 Feldspieler aufgeboden, für jene von 2017 waren es ebenfalls 37 Akteure (Hockey Reference, 2018). Es wird davon ausgegangen, dass sich eine Nomination positiv auf das Salär auswirkt. Diese boolesche Variable wurde bei allen nominierten Spielern mit dem Wert 1 ergänzt, auch wenn sie verletzungsbedingt nicht teilnehmen konnten. Nicht nominierte Spieler weisen den Wert 0 auf. Die Kennzahl DRAFT indiziert, in welcher Runde der jeweilige Spieler gedraftet wurde. Wird ein Spieler früh im Draft von einer NHL-Franchise ausgewählt, sind seine Fähigkeiten als höher einzustufen. Somit dürfte sich ein höherer DRAFT-Wert negativ auf den Lohn auswirken. Das Datenset beinhaltet 217 Einträge von Spielern, die nie gedraftet wurden. Um diese Werte dennoch in der Regression angemessen berücksichtigen zu können, wurden jene Datensätze mit einer fiktiven 10. Drafrunde ergänzt. SOP sagt aus, ob der Spieler nach links oder nach rechts ausgelegt spielt. Spielen mehr Spieler auf links oder auf rechts, könnte dies Auswirkungen auf das Salär haben. Der Datensatz derjenigen Spieler, deren Schusshand links ist, wurde mit dem Wert 0 ergänzt, jene die nach rechts ausgelegt spielen mit dem Wert 1.

¹⁰ Handgeld für den Spieler bei Vertragsunterzeichnung, bezahlt durch das NHL-Team

3.2 Torhüter

Die Leistungskennzahlen der Regular Season 2016/2017 und 2017/2018 sowie die Lohn-
daten aller NHL-Torhütern stammen von Vollman und Cane (2018). Hockey Reference
(2018) bereitete sämtliche Statistiken für die Playoffs auf. Die Statistik umfasst 184 Da-
tensätze. Anders als bei den Feldspielern gibt es keine Duplikate aufgrund von Transfers
während der Saison. Um die Korrektheit der Daten zu verifizieren wurden drei Datensätze
des Datensets mit den offiziellen Statistiken der NHL verglichen.

Da in allen bisherigen Forschungen die Torhüter nicht berücksichtigt wurden, gibt es
keine Anhaltspunkte über signifikante und insignifikante Variablen, welche die Lohnun-
terschiede erklären. Aus diesem Grund beinhaltet der zu analysierende Datensatz die am
häufigsten geführten Statistiken über NHL-Torhüter. In der folgenden Tabelle ist analog
zu den Daten für Feldspieler zwischen allgemeinen Daten zur Person im oberen Ab-
schnitt, Leistungskennzahlen aus der Regular Season im Mittelteil und Statistiken zum
Playoff im unteren Drittel zu unterscheiden. Für die Variablen PPSA, PPGA, PPSP,
SHSA, SHGA und SHSP sind keine Playoff-Statistiken vorhanden.

Tabelle 2: Analyzierte Variablen für Torhüter

Variable	Bezeichnung
SAL	Jahreslohn in Mio. US-Dollar (Cap Hit)
LNSAL	Natürlicher Logarithmus des Jahreslohns in Mio. US-Dollar (Cap Hit)
ID	Spieler-ID
PLAYER	Vorname und Name
SEASON	Saison, aus welcher die analysierten Leistungskennzahlen stammen
AGE	Alter in ganzen Jahren am Ende der jeweiligen Saison (SEASON)
HEIGHT	Körpergröße in Zentimeter
WEIGHT	Körpergewicht in Kilogramm
STAR	Boolesche Variable; Wahr, falls Teilnahme an den All-Star Games 2017/2018
DRAFT	Runde, in welcher der Spieler gedraftet wurde
TEAM	NHL-Team, das den Spieler unter Vertrag hat
POS	Spielerposition
GP	Absolvierte Spiele in der Regular Season
W	Gewonnene Spiele in der Regular Season
L	Verlorene Spiele in der Regular Season
OTL	Niederlagen nach Verlängerung oder Penaltyschiessen in der Regular Season
SA	Erhaltene Torschüsse in der Regular Season

SV	Gehaltene Torschüsse in der Regular Season
GA	Zugelassene Gegentore in der Regular Season
SPE	Fangquote in Prozent in der Regular Season
GAA	Durchschnittlich zugelassene Gegentore pro Spiel in der Regular Season
GSAA	Zugelassene Gegentore über dem Durchschnitt in der Regular Season
MIN	Gespielte Minuten (Karrieretotal) in der Regular Season
SO	Absolvierte Spiele ohne Gegentor (Shutouts) in der Regular Season
G	Erzielte Tore in der Regular Season
A	Erzielte Assists in der Regular Season
PTS	Erzielte Punkte (Tore und Assists) in der Regular Season
PIM	Anzahl Strafminuten in der Regular Season
PPSA	Erhaltene Torschüsse bei gegnerischem Powerplay in der Regular Season
PPGA	Zugelassene Gegentore bei gegnerischem Powerplay in der Regular Season
PPSP	Fangquote in Prozent bei gegnerischem Powerplay in der Regular Season
SHSA	Erhaltene Torschüsse bei gegnerischem Boxplay in der Regular Season
SHGA	Zugelassene Gegentore bei gegnerischem Boxplay in der Regular Season
SHSP	Fangquote in Prozent bei gegnerischem Boxplay in der Regular Season
<hr/>	
PGP	Absolvierte Spiele in den Playoffs
PW	Gewonnen Spiele in den Playoffs
PL	Verlorene Spiele in den Playoffs
POTL	Niederlagen nach Verlängerung oder Penaltyschiessen in den Playoffs
PSA	Erhaltene Torschüsse in den Playoffs
PSV	Gehaltene Torschüsse in den Playoffs
PGA	Zugelassene Gegentore in den Playoffs
PSPE	Fangquote in Prozent in den Playoffs
PGAA	Durchschnittlich zugelassene Gegentore pro Spiel in den Playoffs
PGSAA	Zugelassene Gegentore über dem Durchschnitt in den Playoffs
PMIN	Gespielte Minuten (Karrieretotal) in den Playoffs
PSO	Absolvierte Spiele ohne Gegentore (Shutouts) in den Playoffs
GP	Erzielte Tore in den Playoffs
PA	Erzielte Assists in den Playoffs
PPTS	Erzielte Punkte (Tore und Assists) in den Playoffs
PPIM	Strafminuten in den Playoffs
<hr/>	

Analog zu den Feldspielern werden auch in der Analyse der Torhüter die logarithmierten Saläre des Cap Hits (LNSAL) als abhängige Variable analysiert. Konträr zur Bedeutung für Feldspieler ist davon auszugehen, dass HEIGHT und WEIGHT sich nicht bedeutend

auf das Gehalt auswirken. Eine Nominierung für die All-Star Games (ALLSTAR) und eine hohe Position im Draft (DRAFT) (vgl. 3.1) werden als positive Lohnfaktoren vermutet. Für die All-Star Games 2017 und 2018 wurden jeweils acht Torhüter aufgeboten (Hockey Reference, 2018). Die Datensätze für nicht gedraftete Torhüter wurden mit einer fiktiven 10. Drafrunde ergänzt. Tabelle 2 zeigt, dass sich die Statistik über Leistungskennzahlen der Torhüter fundamental von jener für Feldspieler unterscheidet. Die Variablen GP, W oder SP lassen vermuten, dass den Torhütern eine Schlüsselrolle im Team zukommt. W beinhaltet Partien, die entweder in der regulären Spielzeit, in der Verlängerung oder im Penaltyschiessen gewonnen wurden. Die Leistungskennzahlen SO und PSO dürften sich nicht gross auf das Salär auswirken, da sie für den Teamerfolg eine untergeordnete Rolle spielen und deren Einfluss zu grossen Teilen bereits in der Variable W widerspiegelt sind. Wenn ein Torhüter kein Tor zulässt, gewinnt seine Mannschaft das Spiel in den meisten Fällen. Die Variable GSAA wird von Balloch (2014) wie folgt definiert:

”You take the league’s average save percentage [SP] and apply it to the amount of shots a particular goalie has faced [SA]. You get a number of goals that the average goalie in that league would have surrendered if they faced the same number of shots as the goaltender in question. That number gets compared to the number of goals surrendered by that goaltender, and a plus/minus is created. If a goalie is in the positive, that is how many goals they have saved compared to a league-average goalie. If they are in the negative, then it is safe to assume that they are performing worse than how a league-average goaltender would perform in the same situation”.

Ein Vorteil von GSAA liegt laut Balloch (2014) in seiner guten Vergleichbarkeit. Ein Torhüter mit einer minim schlechteren Fangquote (SP), der aber vielen Schüssen gegen sein Tor (GA) ausgesetzt ist, ist als besser einzustufen als ein Kontrahent, welcher zwar eine bessere Fangquote aufweist, jedoch viel weniger Schüssen ausgesetzt war. Ersterer hat in absoluten Zahlen mehr Chancen des Gegners vereitelt als letzterer. Ein Nachteil liegt darin, dass erhaltene Tore im Boxplay und die Einflüsse auf die Fangquote (SP) mitberücksichtigt werden. Da Mannschaften unterschiedlich vielen Boxplays ausgesetzt sind, wird die Vergleichbarkeit von GSAA beeinträchtigt.

4 Methodik

Die Analyse der in Kapitel 3 beschriebenen Datensets und somit die Beantwortung der Forschungsfrage erfolgt durch die Modellierung von einfachen und multiplen linearen Regressionen. Dieses Kapitel erläutert die Theorie über Regressionen und ihre Anwendung auf Leistungskennzahlen und Löhnen von NHL-Spielern.

4.1 Theoretische Grundlagen der multiplen linearen Regression

Die Regressionsanalyse ist eine statistische Technik, um das Verhalten von Variablen zu untersuchen und zu modellieren (Montgomery, Peck & Vining, 2012, S. 1). Regressionen werden unter anderem für Datenbeschreibungen, Parameterschätzungen, Vorhersagen und Kontrollen eingesetzt (Montgomery et al., 2012, S. 9).

Eine multiple lineare Regression erklärt den Wert einer abhängigen Variable aufgrund des Verhaltens von mehreren unabhängigen Variablen durch eine Funktion in einem mehrdimensionalen Koordinatensystem (Montgomery et al., 2012, S. 68). Allgemein unterscheidet man zwischen Populationsregressionen (PRF) und Stichprobenregressionen (SRF). Laut Newbold, Carlson und Thorne (2013, S. 423) analysiert eine Populationsregression alle Daten einer Grundgesamtheit. Die mathematische Notation einer multiplen linearen Populationsregression lautet:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

In der obigen, von Montgomery et al. (2012, S. 68) beschriebenen PRF wird die abhängige Variable, genannt Regressand, mit y gekennzeichnet. Die unabhängigen Variablen, genannt Regressoren, werden mit x_k notiert. Jeder Regressor hat einen Regressionskoeffizienten, symbolisiert mit β_k . Der Regressionskoeffizient β_k beschreibt die erwartete Veränderung von y bei einer Veränderung von x_k um eine Einheit, unter der Voraussetzung, dass alle anderen Regressoren unverändert bleiben. β_0 bestimmt welche Punkt auf der vertikalen y -Achse die Regression schneidet (Newbold et al., 2013, S. 419). β_0 ist identisch mit y wenn sämtliche $x_k = 0$ sind.

Die Differenz zwischen dem tatsächlich observierten Wert der Variable y_i und dem Wert der Regressionsfunktion ($\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$) wird durch ε symbolisiert. Dieser Errorterm umfasst alle Einflüsse, die durch die Regression nicht erklärt werden (Montgomery et al., 2012, S. 2).

Die Stichprobenregression (SRF) (Newbold et al., 2013, S. 419) hat dieselbe Funktionsweise wie eine Populationsregression, allerdings analysiert sie nur einen Teil der Grundgesamtheit, genannt Stichprobe:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + \varepsilon \quad (2)$$

Bei einer Stichprobenregression wird y als \hat{y} notiert und b_k wird als Schätzer von β_k verwendet. Die unabhängigen Variablen x_k behalten ihre Notation. Der Errorterm ε wird in Stichprobenregressionen als Residuum bezeichnet, seine Funktion bleibt gleich. Oftmals wird versucht, aus der Stichprobenregression auf die Grundgesamtheit zu schliessen (Newbold et al., 2013, S. 419 – 422).

Nachdem das Regressionsmodell spezifiziert wurde, kann man die Stichprobenparameter b_k mithilfe des gesammelten Datensatzes durch die Methode der kleinsten Quadrate (OLS) bestimmen. OLS modelliert die Regression so, dass die kleinstmögliche Summe aller quadrierten Differenzen zwischen den Observationen \hat{y}_i und der Regression ($b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$) gesucht wird (Montgomery et al., 2012, S. 13). Die Methode der kleinsten Quadrate eignet sich nicht für eine Regressionskonstruktion, falls die analysierten Daten nicht normalverteilt sind (Wiseman & Chatterjee, 2010, S. 1). In dieser Arbeit wird die Schätzung der Parameter und alle andere Rechenoptionen mit der Programmiersprache Python gemacht.

Jede Regression verfügt über ein Bestimmtheitsmass R^2 . Es sagt aus, wie hoch die Variabilität in y ist, nachdem sämtliche x bestimmt wurden. Ein hoher Wert für R^2 indiziert, dass die Regression die Veränderungen in y durch Veränderungen der x zu einem grossen Teil erklären kann. Werte für R^2 liegen immer zwischen 0 und 1 (Newbold et al., 2013, S. 433).

Bei einer multiplen Regression zieht man mehrere Regressoren x_k in die Analyse mit ein. Es ist dabei von zentralem Interesse zu wissen, welche Regressoren einen signifikanten Einfluss auf die Veränderung von y haben. Es gilt zu analysieren, ob die zu den Regressoren gehörenden Koeffizienten der Stichprobenregression b_k signifikant genug sind, um in die Populationsregression integriert zu werden. Dies erfolgt durch Hypothesentests der Stichprobenkoeffizienten b_k . Die beiden Hypothesen lauten:

$$H_0 = 0 \quad , \quad H_1 \neq 0 \quad (3)$$

Die Nullhypothese H_0 behauptet, der untersuchte Stichprobenregressionskoeffizient b_k hat keinen signifikanten Einfluss auf die Veränderung von y . Dies bedeutet im Umkehrschluss, dass wenn die Nullhypothese H_0 verworfen werden kann, die Alternativhypothese H_1 angenommen werden muss. Wird die Alternativhypothese angenommen, ist der Regressionskoeffizient signifikant für die Erklärung der Veränderungen von y . Die Nullhypothese wird durch einen t -Test geprüft, welche durch Python ausgeführt wird. Sie gibt für jeden Parameter b_k der Stichprobenregression einen t -Wert aus. Dieser t -Wert wird mit dann mit den Wahrscheinlichkeiten der Standardnormalverteilung in den dazugehörigen p -Wert umgewandelt. Ist der p -Wert kleiner als 0.1, kann H_0 verworfen werden und der getestete Regressionskoeffizient b_k gilt als signifikant (Newbold et al., 2013 S. 438 – 443).

4.2 Analyse von Leistungskennzahlen und Spielerlöhnen

Die in Kapitel 4.1 beschriebene Theorie wird in Kapitel 5 auf Leistungskennzahlen und ihren Einfluss auf die Löhne von NHL-Spielern angewendet. Dabei ist \hat{y} in allen modellierten Regressionen das Salär. Ausgewählte Leistungskennzahlen bilden die Regressoren x_k . Die Stichprobe umfasst Leistungskennzahlen sämtlicher NHL-Spieler aus der Saison 2016/2017 und 2017/2018, die mindestens zehn Spiele in der Regular Season absolviert haben.

Gemäss Kapitel 4.1 eignet sich eine Regression nach OLS nicht, wenn die zugrundeliegenden Daten nicht normalverteilt sind. Kapitel 5.1 identifiziert aus den in Kapitel 3 beschriebenen 52 möglichen Regressoren für Feldspieler und deren 43 für Torhüter die je zehn relevantesten. Als Basis dazu dienen die Erkenntnisse aus dem Literatur Review und der sportökonomische Sachverstand. Ihr Verhalten wird mit Histogrammen (5.1.2.1) und linearen Einfachregressionen gegenüber dem logarithmierten Lohn der Spieler (5.1.2.2 – 5.1.2.4) untersucht und diskutiert. Abschnitt 5.1.3 gibt einen Überblick über die Mittelwerte ausgewählter Variablen. Die Häufigkeitsverteilung der Saläre wird in Kapitel 5.1.4 thematisiert.

Aus Kapitel 2.3 geht hervor, dass die Löhne von professionellen Sportlern keiner Normalverteilung folgen. Aus diesem Grund werden die Gehälter des analysierten Datensatzes logarithmiert. Da eine zu hohe Anzahl an Regressoren das Bestimmtheitsmass R^2 der Regression unter anderem aufgrund von Autokorrelationen verfälschen kann (Newbold et al., 2013, S. 582) und die Anzahl der Freiheitsgrade abnimmt (Newbold et al., 2013, S.

440), beschränken sich die Regressionen auf zehn Variablen. Die Regressionen haben demnach folgende Form:

$$\ln(\hat{y}) = b_0 + b_1x_1 + b_2x_2 + \dots + b_{10}x_{10} + \varepsilon \quad (4)$$

Durch das Logarithmieren von (\hat{y}) beschreibt die Regression, um wie viel Prozent sich (\hat{y}) , also das Salär, verändert, wenn sich die Leistungskennzahl x_k um eine Einheit verändert, *ceteris paribus* (Newbold, Carlson, & Thorne, 2013, S. 517-518). Die konkrete Prozentzahl lässt sich aus dem Koeffizienten b_k ablesen. Die Höhe der Koeffizienten der signifikanten unabhängigen Variablen beantworten die Forschungsfrage dieser Arbeit.

Wie in Kapitel 4.1 bereits erwähnt, modelliert der Computer sämtliche in dieser Arbeit beschriebenen und interpretierten Regressionen. Die multiplen linearen Regressionen werden mit der Programmiersprache Python konstruiert. Alle Codes und Regressionsresultate, inklusive der Ergebnisse der *t*-Tests und den damit zusammenhängenden *p*-Werten, können aus dem Anhang 3 entnommen werden. Der *y*-Achsenabschnitt b_0 wird von Python in den Regressionen mit «const» gekennzeichnet.

Einflussfaktoren, die sich negativ auf das Bestimmtheitsmass R^2 auswirken und somit den Errorterm (ε) der Regression grösser werden lassen, sind einerseits nicht quantifizierbare Spielereigenschaften wie Führungsqualitäten oder das Verhandlungsgeschick bei Lohngesprächen. Ein weiterer möglicher Einfluss ist die in Kapitel 2.2 erwähnte Lohndiskrimination (vgl. auch 6.1).

5 Datenanalyse

Dieses Kapitel analysiert die in Kapitel 3 beschriebenen Leistungskennzahlen in Bezug auf ihren Einfluss auf das Salär. Analysiert wurden Datensätze von Spielern, die zehn oder mehr Spiele in der jeweiligen Regular Season absolviert haben.

Kapitel 5.1 identifiziert die jeweils zehn wichtigsten Leistungskennzahlen für die Regressionskonstruktionen, basierend auf dem Literatur Review. Es zeigt ihr Verhalten anhand von Histogrammen und linearen Einfachregressionen auf und berechnet die Mittelwerte ausgewählter Variablen. Weiter wird die Häufigkeitsverteilung der Variablen SAL und LNSAL erläutert und die daraus entstehenden Implikationen für die Regressionskonstruktion erklärt. In Abschnitt 5.2 werden die Regressionen nach Spielerpositionen modelliert. Die im Literatur Review erwähnten Spielertypen (vgl. 2.3) dienen als Grundlage

für die Regressionen in Kapitel 5.3. Es gilt zu untersuchen, ob Regressionen nach Spielertypen bessere Bestimmtheitsmasse und mehr signifikante Einflussfaktoren auf die Säläre liefern, als Regressionen nach Spielerpositionen. Kapitel 5.4 diskutiert die Regressionsresultate und beantwortet die Forschungsfrage.

5.1 Relevante Leistungskennzahlen

Kapitel 3 beschreibt 52 Variablen für Feldspieler und 42 für Torhüter. In Bezug auf die in Kapitel 4.2 erläuterten Gründe der Autokorrelation von Leistungskennzahlen werden die zehn relevantesten Leistungskennzahlen für Feldspieler und Torhüter identifiziert. Dies geschieht durch eine Analyse des Literatur Reviews und der Anwendung des sportökonomischen Sachverstandes (5.1.1). Das Verhalten der als relevant eingestuften Leistungsvariablen wird danach mit Histogrammen und linearen Einfachregressionen untersucht (5.1.2). Einige Durchschnittswerte der als Regressoren qualifizierten Variablen in Kapitel 5.1.3 geben einen abschliessenden Überblick über die Datensätze.

5.1.1 Identifikation der Regressoren

Kapitel 5.1.1.1 fasst die Selektion bisheriger Leistungskennzahlen in der Lohnforschung über Feldspieler in der NHL zusammen. Ergänzt wird die Auswahl der Regressoren durch ausgewählte weitere Variablen. Abschnitt 5.1.1.2 befasst sich mit der Auswahl der Regressoren für die Torhüterposition. Da es bislang keine Anhaltspunkte aus der Forschung gibt, werden die Variablen zur Regressionskonstruktion mithilfe des sportökonomischen Sachverstandes ausgewählt.

5.1.1.1 Feldspieler

Vincent und Eastman (2009b) bezeichneten in ihrer Forschung bezüglich Lohneinfluss unter anderem die Variablen GP, DRAFT, WEIGHT und PTS als signifikant. Allerdings definierten sie die Variable PTS als die durchschnittliche Anzahl Skorerpunkte pro Spiel während der gesamten Karriere. In dieser Arbeit gibt die Variable Aufschluss über die Anzahl Skorerpunkte während der Saison 2016/2017 und 2017/2018, ohne Statistiken über die Playoffs. Idson und Kahane (2000) untersuchten unter anderem die Einflüsse von PTS, PLUSMINUS und PIM. Jones und Walsh (1988) modellierten eine Regression, die unter anderem die unabhängigen Variablen GP, PTS, DRAFT, PIM und WEIGHT beinhaltet. Eastman (1981) untersuchte den Einfluss von AGE, DRAFT und WEIGHT

auf die Spielersaläre. Aufgrund bisheriger Forschungen werden die Variablen AGE, WEIGHT, DRAFT, GP, PTS, PLUSMINUS und PIM in die Regressionen aufgenommen. Da nur 38.53 Prozent der Spieler Rechtsausleger sind, scheint es aufgrund des sportökonomischen Sachverstandes angemessen, den Einfluss der Variable SOP zu untersuchen, auch wenn diese in der bisherigen Forschung nicht berücksichtigt wurde. Weiter wird auch die Variable HES untersucht, da das Körperspiel je nach Position eine wichtige Rolle einnehmen kann. Als einzige Effizienzkenzahl wird die Variable SP analysiert. Es kann davon ausgegangen werden, dass Spieler, die wenige Chancen brauchen, um ein Tor zu erzielen, wichtig für ein Team sind und somit besser bezahlt werden. Ob dieser Einfluss signifikant ist, sollen die Regressionen zeigen. Zusammengefasst qualifizieren sich somit folgende Variablen für die Regressionskonstruktion für Feldspieler.

Tabelle 3: Regressoren für die Lohnanalyse von Feldspielern

Variable	x_k	Bezeichnung
AGE	x_1	Alter in ganzen Jahren am Ende der jeweiligen Saison (SEASON)
WEIGHT	x_2	Körpergewicht in Kilogramm
DRAFT	x_2	Runde, in welcher der Spieler gedraftet wurde
SOP	x_4	Schusshand des Spielers, entweder links oder rechts
GP	x_5	Absolvierte Spiele in der Regular Season
PTS	x_6	Erzielte Punkte (Tore und Assists) in der Regular Season
PLUSMINUS	x_7	Plus/Minus-Statistik in der Regular Season
PIM	x_8	Strafminuten in der Regular Season
SP	x_9	Erfolgreiche Torschüsse in Prozent in der Regular Season
HES	x_{10}	Ausgeführte Checks bei numerischer Gleichheit in der Regular Season

Die in der bisherigen Forschung oftmals analysierte Variable STAR wird aus Gründen der starken Autokorrelation mit der Leistungskennzahl DRAFT nicht in die Regression miteinbezogen. Eine Auswertung des Datensatzes ergab, dass für die Saisons 2016/2017 und 2017/2018 insgesamt nur 20 Spieler eine Einladung zu All-Star Games erhielten, die nicht in den ersten zwei Runden gedraftet wurden. Insgesamt wurden in den beiden Jahren 90 Spieler für die All-Star Games nominiert. Leistungskennzahlen aus dem Playoff wurden gänzlich ausgeschlossen, da viele Spieler kein Playoff spielen. Da die Variable WEIGHT stark mit der Variable HEIGHT korreliert und beide Ähnliches über eine physische Spielweise aussagen, beinhaltet die Regression nur die Variable WEIGHT.

5.1.1.2 Torhüter

Die Auswahl der Regressoren für Torhüter ist zu einem grossen Teil auf die Statistiken von Vollman und Cane (2018) gestützt. Die für Feldspieler als signifikant vermuteten Variablen AGE, GP und DRAFT werden ebenfalls für die Regression der Torhüterposition berücksichtigt. Die mutmasslich wichtigsten Leistungskennzahlen eines Torhüters sind die Anzahl gewonnener Spiele (W) und seine Fangquote (SPE). Die Variablen GSAA und MIN ergänzen die Kennzahl SPE, weshalb sie sich als Regressoren qualifizieren. GSAA bietet eine ausgezeichnete Vergleichbarkeit und vereint mehrere Leistungskennzahlen. Ein Torhüter steht im Mittelpunkt, wenn sein Team unter Druck ist. Dies ist sehr oft der Fall, wenn seine Mannschaft einem Boxplay ausgesetzt ist. Darum beinhaltet die Regression die Statistiken über die Fangquote in Prozent bei gegnerischem Powerplay in der Regular Season (PPSP). Der Regressor SO soll die in Kapitel 3 behauptete Insignifikanz belegen. Da grössere Torhüter mehr Fläche abdecken, kann einen Zusammenhang zwischen Erfolg und Körpergrösse vermutet werden, weshalb HEIGHT sich als Regressor qualifiziert. Zusammengefasst beinhaltet die Regressionsfunktion für Torhüter folgende Variablen.

Tabelle 4: Regressoren für die Lohnanalyse von Torhütern

Variable	x_k	Bezeichnung
AGE	x_1	Alter in ganzen Jahren am Ende der jeweiligen Saison (SEASON)
HEIGHT	x_2	Körpergrösse in Zentimeter
DRAFT	x_2	Runde, in welcher der Spieler gedraftet wurde
GP	x_4	Absolvierte Spiele in der Regular Season
W	x_5	Gewonnene Spiele in der Regular Season
SPE	x_6	Fangquote in Prozent in der Regular Season
GSAA	x_7	Zugelassene Gegentore Tore über dem Durchschnitt in der Regular Season
MIN	x_8	Gespielte Minuten (Karrieretotal) in der Regular Season
SO	x_9	Absolvierte Spiele ohne Gegentore (Shutouts) in der Regular Season
PPSP	x_{10}	Fangquote in Prozent bei gegnerischem Powerplay in der Regular Season

5.1.2 Histogramme & lineare Einfachregressionen der Regressoren

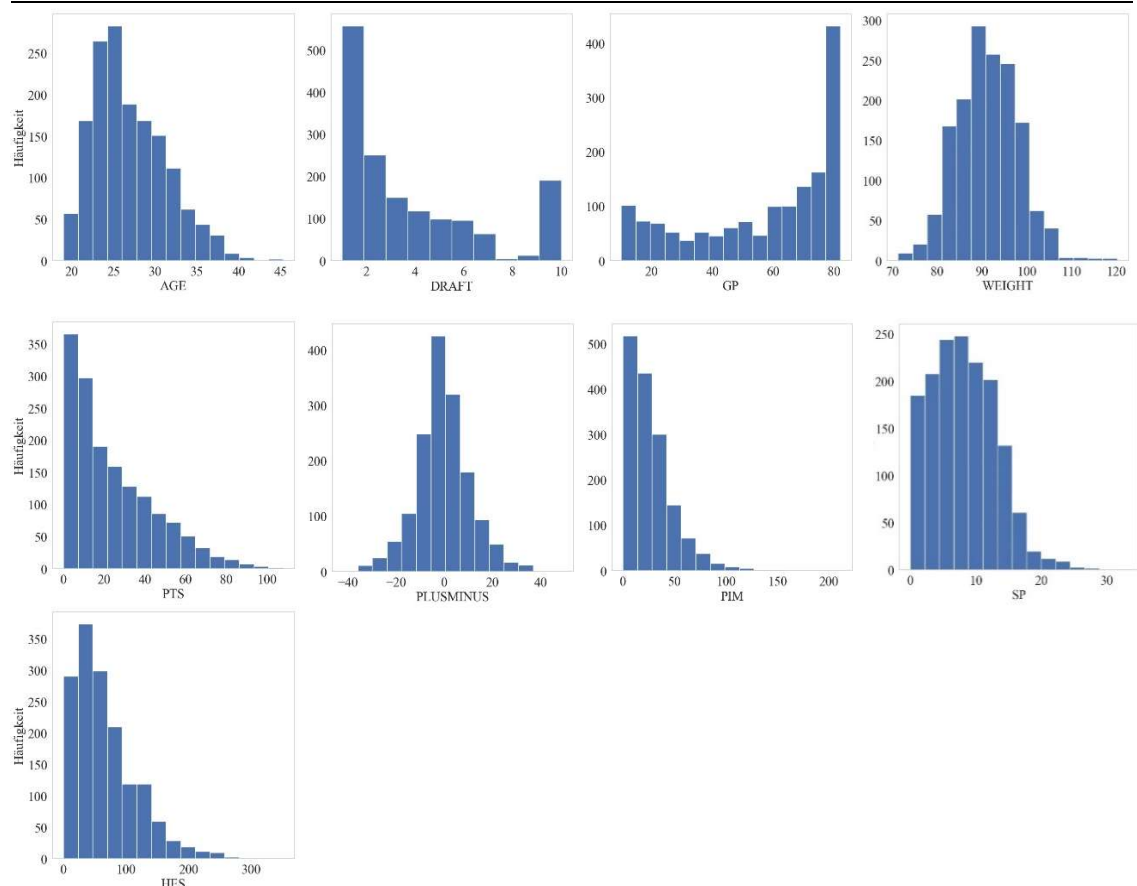
Kapitel 5.1.2.1 bildet die Histogramme der Variablen ab, die für die Regressionen verwendet werden. Abschnitt 5.1.2.2 beinhaltet Streudiagramme der Variablen, die sowohl für die Regressionen der Feldspieler, als auch für jene der Torhüter verwendet werden.

Unter 5.1.2.3 befinden sich die Verteilungen der Regressoren für Feldspieler und Abschnitt 5.1.2.4 bildet das Verhalten der analysierten Leistungskennzahlen für Torhüter ab. Alle Regressoren wurden durch die Streudiagramme auf ihr Verhalten bezüglich des logarithmierten Salärs (LNSAL) untersucht und mit einer linearen Regressionsgeraden versehen. Die hellblaue Schattierung um die Regressionsgerade gibt Aufschluss über das Bestimmtheitsmass R^2 . Je kleiner die Fläche, desto grösser das Bestimmtheitsmass. Ob das Bestimmtheitsmass gross genug ist, um die Variable als signifikant zu klassifizieren, wird durch die Regressionsanalysen in Kapitel 5.2 und 5.3 ermittelt.

5.1.2.1 Histogramme

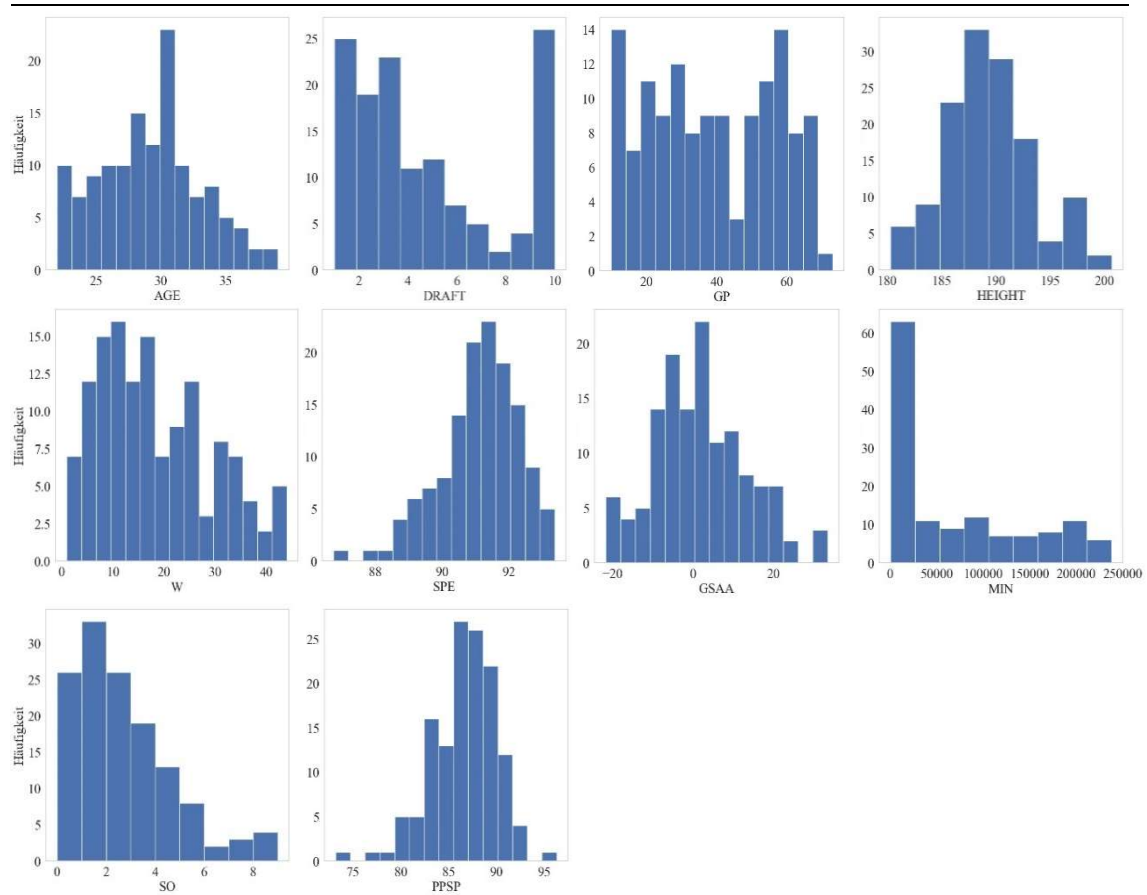
Die in den Histogrammen visualisierten Variablen sind nach Positionen aufgeteilt. Abbildung 12 gibt einen Überblick über die Häufigkeitsverteilungen der Kennzahlen für Feldspieler. Abbildung 13 zeigt das Verhalten der Statistiken für die Torhüterposition. Die Histogramme belegen, dass kein Logarithmieren der unabhängigen Variablen notwendig ist, um eine aussagekräftige Regression zu modellieren, da keine der Statistiken grössere Ausreisser beinhaltet.

Abbildung 1: Histogramme der Regressoren für Feldspieler



Die Variablen AGE, WEIGHT, PLUSMINUS und SP folgen einer Normalverteilung, was aufgrund ihrer Eigenschaften zu erwarten war. Einer rechtsschiefen Verteilung folgen die Variablen PTS, PIM und HES. Konträr dazu weist das Histogramm über die Leistungskennzahl GP eine linksschiefe Verteilung auf. Der Ausreisser am rechten Ende zeigt, dass viele der analysierten Feldspieler beinahe alle Spiele der Regular Season (82) absolviert haben. Die Statistik DRAFT weist eine vergleichsweise ungewöhnliche Verteilung auf, da der Datensatz mit einer fiktiven 10. Drafrunde ergänzt wurde, um ungedraftete Spieler für die Regressionen zu berücksichtigen. Diese Tatsache begründet den Ausreisser am rechten Ende des Histogramms. Aus der Verteilung der Variable DRAFT geht weiter hervor, dass früher gedraftete Spieler bessere Fähigkeiten aufweisen, weshalb sie sich auch eher in der NHL durchsetzen können.

Abbildung 2: Histogramme der Regressoren für Torhüter



Analog zu den Feldspielern ist die Variable AGE auch bei den Torhütern normalverteilt. Gleiches trifft auf die Kennzahl HEIGHT, SPE, GSAA und PPSP zu, wovon auszugehen war. Rechtsschief sind die Statistiken über W und MIN, wobei die Statistik MIN einen sehr grossen Ausreisser am linken Ende des Histogramms aufweist. Dieser lässt darauf

schliessen, dass eine grosse Anzahl Torhüter noch nicht viele Spiele in der NHL bestritten haben, was als Gegensatz zur normalverteilten Altersstatistik interpretiert werden kann. Die Leistungsvariable GP ist für Torhüter weniger stark linksverteilt als für Feldspieler. Laut der allgemeingültigen Meinung kann dies als überraschend aufgefasst werden. Feldspieler tragen normalerweise ein höheres Risiko, Spiele zu verpassen, als Torhüter, beispielsweise durch Verletzungen.

5.1.2.2 Regressoren für Feldspieler & Torhüter

Die Variablen AGE, DRAFT und GP werden sowohl für die Regressionen der Feldspieler als auch für die Regression der Torhüter verwendet, weshalb Abbildung 3 bis 5 die Datensätze aller Spieler in die Regression miteinbeziehen, die mehr als 10 Ligaspiele bestritten haben.

Abbildung 3: Einfache Linearregressionen LNSAL/AGE für alle Spielerpositionen

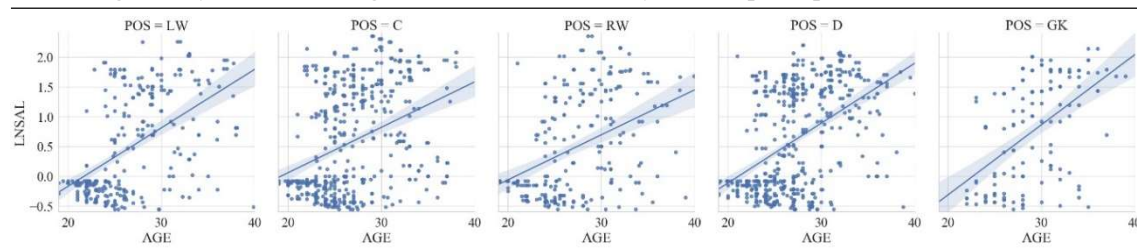


Abbildung 3 zeigt auf, dass ein höheres Alter sich über alle Spielerpositionen hinweg positiv auf das Salär auswirkt. Das Bestimmtheitsmass der Regression für Verteidiger ist am grössten. Den stärksten Lohnanstieg pro Altersjahr erhalten Torhüter.

Abbildung 4: Einfache Linearregressionen LNSAL/DRAFT für alle Spielerpositionen

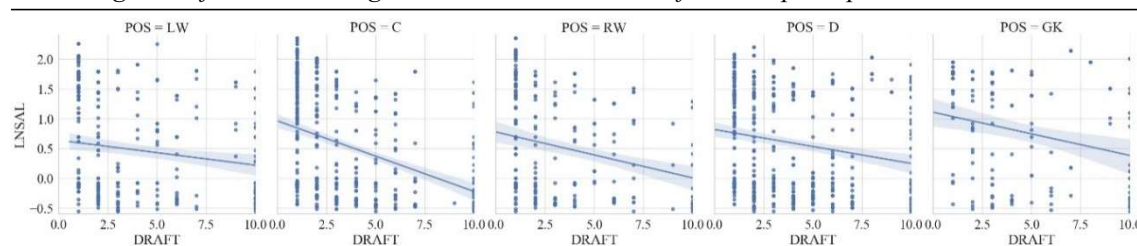
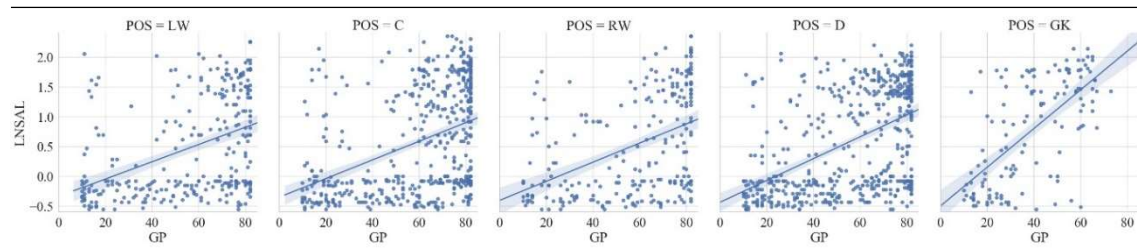


Abbildung 4 zeigt den im Literatur Review erwähnten Zusammenhang zwischen der Position im Entry Draft und der Höhe des Salärs. Je früher ein Spieler im Draft gezogen wird, desto besser werden seine Fähigkeiten eingeschätzt. Der höhere Lohn für früher gedraftete Spieler zeigt auf, dass die Einschätzungen der Fähigkeiten für den Entry Draft

relativ akkurat sind. Es gibt auf der Verteidigerposition mehr ungedraftete Spieler (72) als auf den Stürmerpositionen (LW=42, C=52, RW=25) und der Torhüterposition (26). Ungedraftete Spieler werden auf der 10. fiktiven Drafrunde geplottet.

Abbildung 5: Einfache Linearregressionen LNSAL/GP für alle Spielerpositionen



Die linearen Regressionen in Abbildung 5 bestätigen die angenommene Korrelation zwischen einem höheren Salär für Spieler, die mehr Spiele absolviert haben. Es fällt auf, dass sehr viele Feldspieler beinahe sämtliche 82 Spiele der Regular Season 2016/2017, beziehungsweise 2017/2018 bestritten haben. Konträr dazu steht die Statistik der Torhüter. Cam Talbot von den Edmonton Oilers, absolvierte in der Saison 2016/2017 73 Partien und war damit der einzige Torhüter, der in den zwei analysierten Saisons über 70 Spiele bestritt. Der zusätzliche Lohn pro Spiel für die Position GK indes am höchsten.

5.1.2.3 Regressoren für Feldspieler

Nebst den drei in Kapitel 5.1.2.1 analysierten Kennzahlen AGE, DRAFT und GP beinhalten die Regressionen für Feldspieler die Leistungskennzahlen WEIGHT, PTS, PLUSMINUS, PIM, SP und HES, auf welche im Folgenden genauer eingegangen wird. Eine lineare Regression für den booleschen Regressor SOP bietet keinen Mehrwert, weshalb auf eine Visualisierung verzichtet wird.

Abbildung 6: Einfache Linearregressionen LNSAL/WEIGHT für Feldspieler

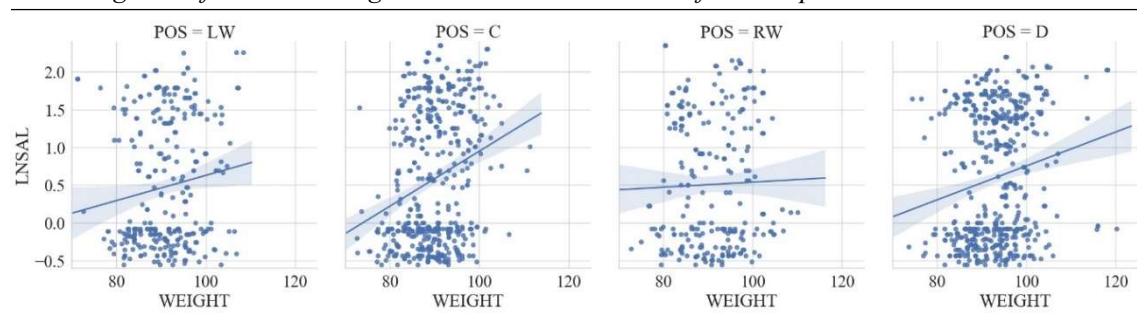


Abbildung 6 zeigt eine positive, wenn auch nicht allzu eindeutige Korrelation zwischen LNSAL und WEIGHT. Der Einfluss des Gewichts auf den Lohn ist für Center und Verteidiger deutlich stärker als für RW und LW. Die unterschiedliche Länge der Regressionsgeraden zeigt, dass der schwerste Verteidiger (Nikita Tryamkin, 120.2 Kilogramm) deutlich schwerer ist, als der schwerste Stürmer (Brian Boyle, 111.36 Kilogramm).

Abbildung 7: Einfache Linearregressionen LNSAL/PTS für Feldspieler

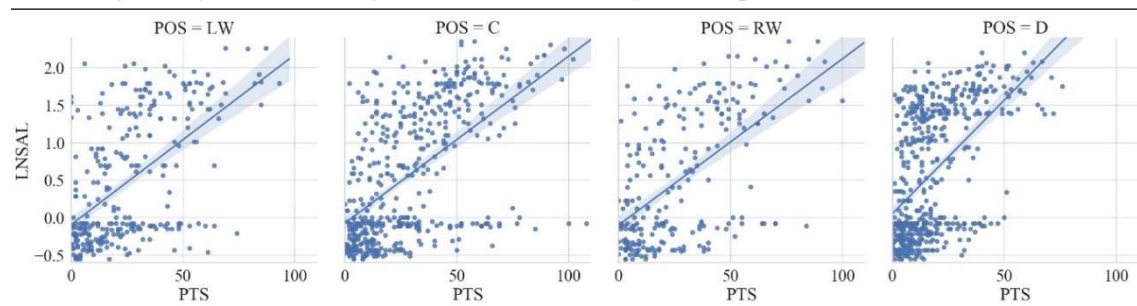
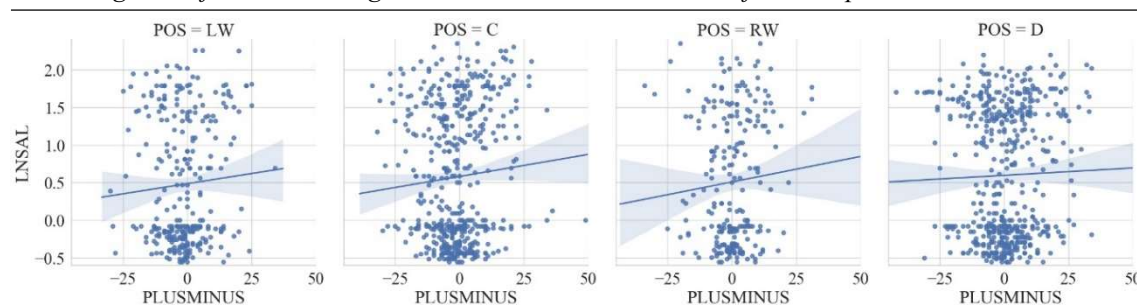


Abbildung 7 bestätigt den angenommenen starken Zusammenhang zwischen erzielten Skorerpunkten und Lohn. Des Weiteren ist anhand der unterschiedlichen Streuung festzustellen, dass Stürmer mehr Skorerpunkte erzielen als Verteidiger. Das Maximum liegt bei 108, erreicht von Connor McDavid, dem Mittelstürmer der Edmonton Oilers, in der Saison 2017/2018. Den Bestwert eines Verteidigers betrug 76 Punkte, erreicht von Brent Burns, Verteidiger der San Jose Sharks in der Saison 2016/2017.

Abbildung 8: Einfache Linearregressionen LNSAL/PLUSMINUS für Feldspieler



Eine positive Plus/Minus-Bilanz wirkt sich gemäss den Regressionsgeraden in Abbildung 8 im Mittel positiv auf den Lohn aus. Die Regressionsgüte ist jedoch unbefriedigend. Die Spannweite der Plus/Minus-Bilanzen für linke Flügelstürmer ist mit 64 am kleinsten, jene für Center mit 82 am grössten. Über die beste Plus/Minus-Bilanz verfügte der Mittelstürmer der Vegas Golden Knights, William Karlsson, in der Saison 2017/2018, mit +49 bei 82 gespielten Partien. Die geringe Regressionsgüte und die damit zusammenhängende

Ambiguität der Plus/Minus-Statistik in Bezug auf die Auswirkungen auf das Salär ist insbesondere bei den Verteidigern stark ausgeprägt. Ryan Suter von den Minnesota Wild wies in der Saison 2016/2017 eine Plus/Minus-Bilanz von +24 bei einem Salär von 7.538 Mio. US-Dollar auf. Den gleichen Wert erspielte sich auch Josh Manson von den Anaheim Ducks in der darauffolgenden Saison. Allerdings war sein Jahresgehalt mit 825'000 US-Dollar deutlich tiefer. Nick Leddy von den New York Islanders verdiente in der Saison 2017/2018 5.5 Mio. US-Dollar. Er erreichte dabei eine schlechte Plus/Minus-Bilanz von -42. Die Bilanz von Damon Severson von den New Jersey Devils wies eine Saison zuvor einen Wert von -31 bei einem jährlichen Lohn von 606'000 US-Dollar auf. Diese Beispiele ähnlicher Plus/Minus-Statistiken bei ungleichem Gehalt belegen die Ambiguität der Variable PLUSMINUS.

Abbildung 9: Einfache Linearregressionen LNSAL/PIM für Feldspieler

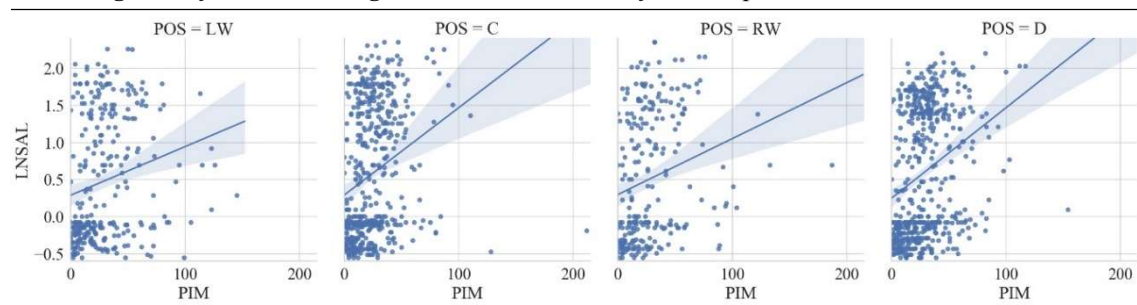


Abbildung 9 zeigt eine überraschend positive Korrelation des Gehalts mit der Anzahl Strafminuten (PIM). Dieser Zusammenhang könnte aber auch darauf zurückgeführt werden, dass Spieler, die mehr Spiele absolvieren, auch mehr Strafminuten aufweisen. Da der Zusammenhang von GP und LNSAL gemäss Abbildung 3 stark positiv korreliert, könnte dies die ebenfalls positive Korrelation von PIM und LNSAL erklären. Zudem ist die Güte dieser linearen Regressionen tief.

Abbildung 10: Einfache Linearregressionen LNSAL/SP für Feldspieler

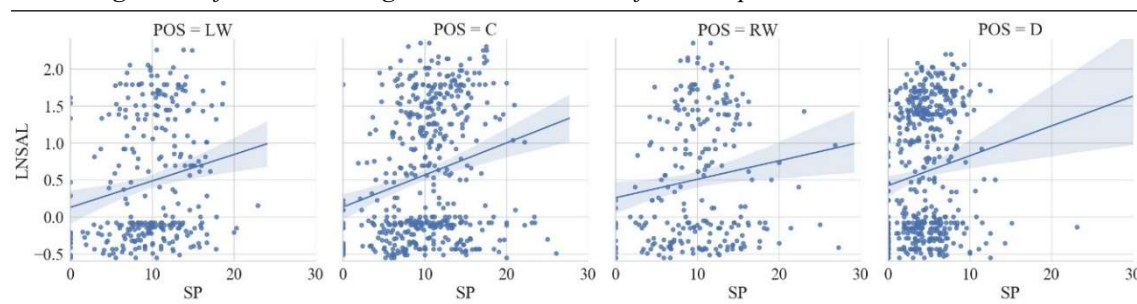
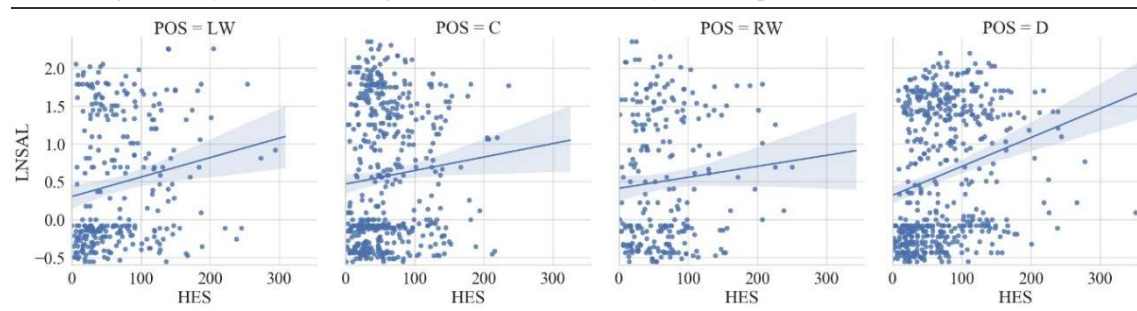


Abbildung 10 zeigt einen positiven, wenn auch schwachen Zusammenhang zwischen einer guten Schussquote und einem hohen Salär. Spieler, die weniger Torschüsse für dieselbe Anzahl Tore brauchen, oder mit gleich vielen Torschüssen mehr Tore erzielen sind produktiver und somit mehr von höherem Wert. Die Korrelation ist für Stürmer deutlich stärker als für Verteidiger.

Abbildung 11: Einfache Linearregressionen LNSAL/HES für Feldspieler



Die Anzahl ausgeführter Checks korreliert gemäss Abbildung 11 stark positiv mit der Höhe des Salärs für Verteidiger. Bei Offensivspielern fällt der Zusammenhang schwächer aus. Über alle Positionen hinweg gesehen ist das Bestimmtheitsmass dieser linearen Regressionen eher gering. Am meisten Checks führte Mark Borowiecki in der Saison 2016/2017 aus. Er checkte seine Gegenspieler in 70 Spielen insgesamt 351 Mal. Das sind durchschnittlich fünf Checks pro Partie.

5.1.2.4 Regressoren für Torhüter

Nebst den drei in Kapitel 5.1.2.1 analysierten Kennzahlen AGE, DRAFT und GP beinhalten die Regressionen für Torhüter die Leistungskennzahlen HEIGHT, W, SPE, GSAA, MIN, SO und PPSP, deren Verhalten nachfolgend analysiert wird.

Abbildung 12: Einfache Linearregressionen LNSAL/HEIGHT, LNSAL/W, LNSAL/SPE & LNSAL/GSAA für Torhüter

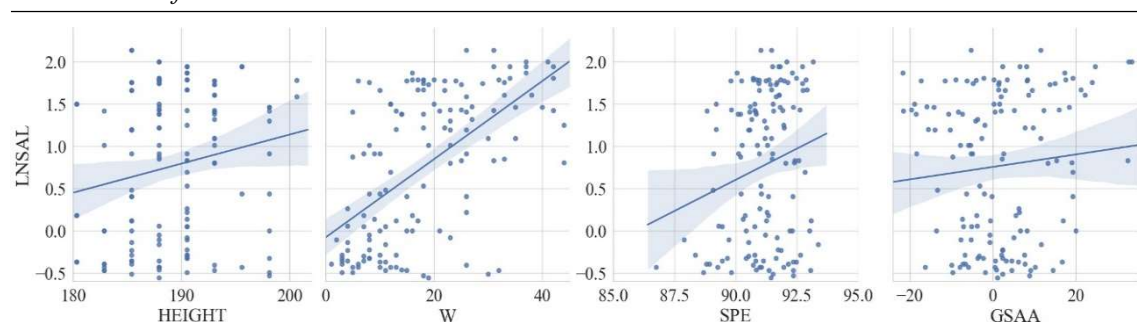
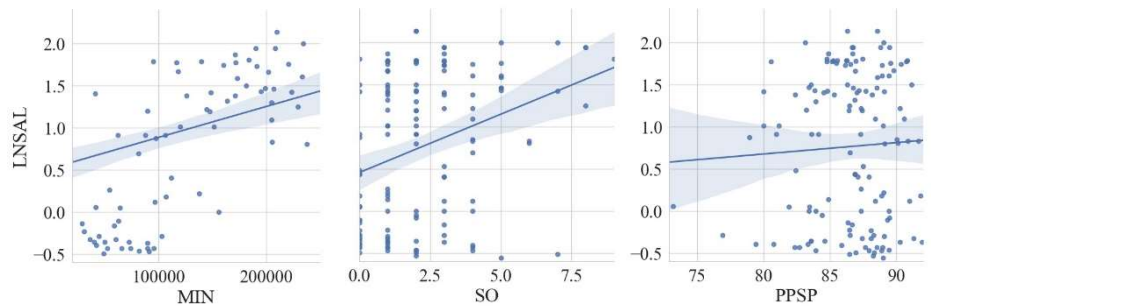


Abbildung 12 bestätigt die starke positive Korrelation zwischen W und $LNSAL$. Weniger deutlich ist der positive Zusammenhang zwischen $HEIGHT$ und $LNSAL$, sowie SPE und $LNSAL$. Keine Aussage über die Auswirkungen auf das Gehalt lässt die Variable $GSAA$ zu, da die Regressionsgüte ungenügend ist.

Abbildung 13: Einfache Linearregressionen $LNSAL/MIN$, $LNSAL/SO$, $LNSAL/PPSP$ für Torhüter



Die Anzahl gespielter Karriereminuten in der Regular Season (MIN) korrelieren positiv mit dem Salär, wie Abbildung 13 zeigt. Dieser Zusammenhang könnte zu einem gewissen Grad auch mit der Korrelation des Alters und des Lohnes erklärt werden, wobei ein höheres Alter nicht zwingend mehr gespielte Minuten impliziert. Entgegen der Erwartungen korreliert die Anzahl Shutouts (SO) positiv mit der Höhe des Lohnes. Über den Zusammenhang von $LNSAL$ und $PPSP$ lässt die lineare Regression keine Schlüsse zu.

5.1.3 Arithmetische Mittelwerte ausgewählter Variablen

Um eine detailliertere Übersicht über die erhobenen Daten zu erhalten, zeigt Tabelle 5 Durchschnittswerte ausgewählter Variablen, aufgeteilt nach Spielerpositionen. Die erste Spalte zeigt den gewichteten Durchschnitt für alle Spalten an, für jene Werte verfügbar sind. Die Zeile SOP zeigt die Anteile der Rechtsausleger. Für die Variablen $DRAFT$ und W wurden keine Mittelwerte berechnet. Die Tabelle berücksichtigt nur Daten der Regular Season. Durchschnittswerte für die Playoffs bringen keinen Mehrwert, da 988 von 1681 Datensätze keine Daten aufweisen, weil sich lediglich 16 von 31 Teams für die Playoffs qualifizieren. Der obere Teil der Daten ist für Feldspieler und Torhüter relevant. Der Mittelteil betrifft ausschliesslich Feldspieler und die Variablen im unteren Drittel sind nur für Torhüter von Bedeutung.

Tabelle 5: Durchschnitte & prozentuale Anteile der Variablen nach Position

Variable	Alle	C	RW	LW	D	GK
<i>n</i>	1681	490	240	288	529	134
SAL (in Mio. USD)	2.57	2.63	2.42	2.33	2.59	3.04
LNSAL (in Mio. USD)	0.57	0.58	0.51	0.48	0.60	0.77
AGE	27.15	26.78	27.54	26.61	27.08	29.20
HEIGHT	185.86	184.43	185.04	184.72	187.32	189.25
WEIGHT	91.27	89.68	90.81	91.01	93.00	91.64
GP	55.96	59.22	57.05	56.31	56.53	39.05
SOP	38.53%	34.69%	70.83%	13.54%	41.02%	-
PTS	24.42	29.19	27.54	25.36	18.07	-
PLUSMINUS	-0.36	-1.15	0.15	-1.21	0.60	-
PIM	27.59	24.40	28.53	29.55	29.04	-
SP	8.15%	10.24%	10.13%	9.87%	4.38%	-
HES	66.61	59.45	66.13	68.02	72.70	-
SPE	-	-	-	-	-	91.12%
GSAA	-	-	-	-	-	1.7278
MIN	-	-	-	-	-	67*780
SO	-	-	-	-	-	2.24
PPSP	-	-	-	-	-	86.67%

Von den ursprünglich 2037 gesammelten Datensätzen (1853 Feldspieler, 184 Torhüter) erfüllen insgesamt 1681 (1547 Feldspieler, 134 Torhüter) die Bedingung von mindestens zehn absolvierten Partien in der Regular Season, um für die Regressionen berücksichtigt zu werden. Während eines Spiels stehen gleichzeitig ein Mittelstürmer (C), ein rechter Flügelstürmer (RW), ein linker Flügelstürmer (LW), zwei Verteidiger (D) und ein Torhüter (GK) auf dem Eis. Dies erklärt, warum *n* für Verteidiger höher ist, als für die Stürmerpositionen. Die zahlreichen Beobachtungen für C im Vergleich zu RW und LW könnten darauf zurückgeführt werden, dass Center auch als Flügelstürmer eingesetzt werden können, ihre bevorzugte Position aber jene des Mittelstürmers ist.

Die in Kapitel 3.2 aufgestellte Behauptung, dass Torhüter zentral für den Erfolg eines Teams seien, scheint sich aufgrund des höheren Durchschnittslohns (SAL) zu bewahrheiten. Er ist deutlich höher, als jener von Feldspielern. Diese These könnte allerdings mit der positiven Korrelation zwischen Alter (AGE) und Lohn (SAL) herausgefordert werden, da Torhüter im Schnitt deutlich älter sind. Die durch Abbildung 5 gewonnene Er-

kenntnis, dass Feldspieler mehr Spiele absolvieren als Torhüter wird durch die Mittelwerte der Variable GP belegt. Verteidiger (D) sind im Mittel grösser und schwerer als Offensivspieler (C, LW und RW). Torhüter (GK) sind durchschnittlich am grössten.

Hervorzuheben sind insbesondere die teilweise grossen Unterschiede in den Mittelwerten zwischen RW und LW. Die Lohndifferenz zwischen den beiden Positionen beträgt im Mittel 90'000 US-Dollar und könnte mit signifikanten Abweichungen in den Statistiken AGE, PTS, PLUSMINUS und SP erklärt werden. Bei allen Statistiken, ausser jener der ausgeführten Checks (HES) sind die Mittelwerte für RW höher als für LW. Die These, dass sich viele Strafminuten negativ auf das Salär auswirken, wurde durch die Korrelation in Abbildung 9 widerlegt. Die Variable PIM scheint also nicht für den tieferen Lohn von linken Flügelstürmern verantwortlich zu sein. SOP beschreibt den Anteil der analysierten Spieler n , die rechts ausgelegt spielen. Der hohe prozentuale Anteil an Rechtsausleger auf der rechten Flügelposition (70.83 Prozent) ist auf Vorteile im Spiel zurückzuführen. Rechtsausleger können bei Angriffen über rechts den Puck besser mit ihrem Körper abschirmen. Gegenteiliges Argument trifft auf linke Flügelstürmer zu, von welchen nur 13.54 Prozent nach rechts ausgelegt spielen. Über alle Kategorien der Feldspieler hinweg gesehen spielen jedoch nur 38.53 Prozent nach rechts ausgelegt. Diese ungleiche Verteilung der Schusshand hat möglicherweise Auswirkungen auf das Salär, da die Schusshand in gewissen Spielsituationen, beispielsweise im Powerplay, entscheidend sein kann.

Die arithmetischen Mittelwerte für Torhüter zeigen, dass die Abwehrquote bei gegnerischem Powerplay (PPSP) um 4.45 Prozent schlechter ist als wenn beide Teams gleich viele Spieler auf dem Eis haben (SPE). Dies ist darauf zurückzuführen, dass das dezierte Team unter Druck spielen muss, was offenbar zu mehr Gegentreffer führt. Der tiefe Mittelwert der Variable MIN kann durch den Ausreisser im Histogramm in Abbildung 2 erklärt werden.

5.1.4 Häufigkeitsverteilung der Spielerlöhne

Wiseman und Chatterjee (2010, S. 1) sagen, dass sich eine Regressionsanalyse nach OLS nicht eignet, wenn die zugrundeliegenden Daten (SAL) nicht normalverteilt sind oder Ausreisser beinhalten. Darum muss die abhängige Regressionsvariable SAL auf ihre Verteilung hin überprüft werden.

Abbildung 14: Histogramm SAL

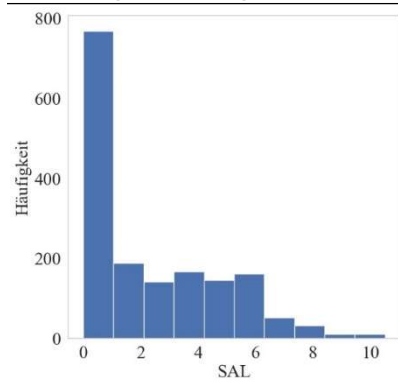
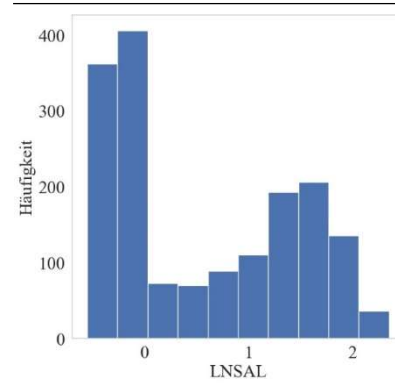


Abbildung 15: Histogramm LNSAL



Das Histogramm in Abbildung 14 zeigt auf, dass die Löhne nicht ansatzweise normalverteilt sind und Ausreisser beinhalten. Abbildung 15 zeigt die Häufigkeitsverteilung der Variable LNSAL. Durch das Logarithmieren der Spielerlöhne wurden Ausreisser im oberen Bereich eliminiert und die Spannweite reduziert. Somit kann eine aussagekräftige Regression nach der OLS-Methode modelliert werden.

5.2 Regressionen nach Spielerpositionen

Die Regressionen nach den Spielerpositionen C, RW, LW, D und GK geben Aufschluss darüber, welche Regressoren einen nennenswerten Einfluss für den Lohn welcher Spielerpositionen haben. Tabelle 6 zeigt für jede Regression die Anzahl Beobachtungen n , die Regressionsgüte R^2 und die p -Werte für jede Leistungskennzahl und jede Position. Signifikante Leistungskennzahlen auf das Salär sind blau eingefärbt. Weiterführende Informationen aus den Regressionen können dem Anhang entnommen werden. Diskutiert werden die Resultate in Kapitel 5.4.

Tabelle 6: p-Werte der Regressionen nach Positionen

Variablen	C	RW	LW	D	GK
<i>n</i>	490	240	288	529	134
<i>R</i> ²	0.570	0.570	0.609	0.546	0.622
AGE	0.000	0.000	0.000	0.000	0.000
GP	0.008	0.022	0.019	0.005	0.010
DRAFT	0.000	0.004	0.013	0.000	0.007
PTS	0.000	0.000	0.000	0.000	-
SOP	0.532	0.548	0.013	0.452	-
PLUSMINUS	0.002	0.197	0.035	0.015	-
PIM	0.605	0.710	0.193	0.721	-
WEIGHT	0.002	0.206	0.072	0.096	-
HES	0.186	0.070	0.233	0.479	-
SP	0.719	0.420	0.066	0.763	-
W	-	-	-	-	0.207
SPE	-	-	-	-	0.909
HEIGHT	-	-	-	-	0.426
GSAA	-	-	-	-	0.993
MIN	-	-	-	-	0.153
PPSP	-	-	-	-	0.462
SO	-	-	-	-	0.095

5.3 Regressionen nach Spielertypen

Um eine Regression nach den in Kapitel 2.4 beschriebenen Spielertypen modellieren zu können, müssen die Datensätze zuerst in die entsprechenden Kategorien (vgl. 2.4) eingeteilt werden. Als Anhaltspunkte zur Kategorisierung dienen dabei die Mittelwerte über Skorerpunkte pro Spiel (PTSPGP), die Plus/Minus-Statistik (PLUSMINUS), die Anzahl Strafminuten pro Spiel und das Körpergewicht (WEIGHT). Um die Einteilung zu vereinfachen, wird der Fokus auf die Variable PTSPG gelegt. Die anderen Variablen dienen dazu, die Einteilung zu verifizieren, indem sie mit den Durchschnittswerten der Arbeit von Vincent and Eastman (2009a) verglichen werden. Die Grundlage für die Kategorisierung sind die Leistungskennzahlen aus der Regular Season 2016/2017 und 2017/2018. Es werden nur Spieler berücksichtigt, die mindestens zehn Spiele bestritten haben.

5.3.1 Kategorisierung Stürmer

Die besten 31 Prozent der Stürmer, gemessen an der Variable PTSPGP, werden als Scorers eingeteilt, die mittleren 57 Prozent als Grinders und die schlechtesten 12 Prozent als Enforcers. Es handelt sich dabei um die die in Kapitel 2.3 erwähnten Proportionen.

Tabelle 7: Kategorisierung der Angreifer nach Spielertypen

Spielertyp	Anteil	n	PTSPGP	PLUSMINUSPGP	PIMPGP	WEIGHT
Scorers	31%	316	0.750	0.045	0.442	90.17
Grinders	57%	580	0.322	-0.041	0.454	90.07
Enforcers	12%	122	0.091	-0.136	0.659	91.92
<i>Total</i>	<i>100%</i>	<i>1018</i>	<i>0.427</i>	<i>-0.015</i>	<i>0.464</i>	<i>90.32</i>

Die Aufteilung der Stürmer in die drei Kategorien Scorers, Grinders und Enforcers ergibt sehr ähnliche Durchschnittswerte, verglichen mit den Ergebnissen von Vincent und Eastman (2009a, S. 9). Die durchschnittlichen Punkte pro Spiel (PTSPGP) und Plus/Minus-Statistik pro Spiel (PLUSMINUSPGP), sind für die Kategorien Scorers und Grinders nahezu identisch. Für Enforcers fallen beide Kennzahlen in dieser Analyse schlechter aus. Die Anzahl Strafminuten pro Spiel (PIMPGP) für die Kategorie Enforcers sind für diesen Datensatz im Mittel deutlich tiefer, als jene von Vincent und Eastman (Vincent & Eastman, 2009a, S. 9) (2.46). Spannend ist, dass der grösste PIM-Wert für Scorers mit 122 Minuten deutlich kleiner ist als jener der «Grinders (187 Minuten) und der Enforcers (212 Minuten). Das durchschnittliche Körpergewicht (WEIGHT) in diesem Datensatz über alle drei Kategorien hinweg gesehen ähnlich. Dies steht im Kontrast zu den Analysen von Vincent und Eastman (2009a), in welchen Enforcers deutlich schwerer waren als Scorers und Grinders. Die Annäherung des Körpergewichts und die deutlich tieferen Werte für PIMPGP könnten mit der Modernisierung des Eishockeysports erklärt werden. Auf eine weitere Analyse bezüglich den Ursachen dieser Veränderungen wird verzichtet. Abschliessend kann gesagt werden, dass die Aufteilung aufgrund der ähnlichen Durchschnittswerte verglichen mit der Arbeit von Vincent und Eastman (2009a), sinnvoll ist und daher als Regressionsgrundlage dienen kann.

5.3.2 Kategorisierung Verteidiger

Die besseren 55 Prozent der Verteidiger, gemessen an der Variable PTSPGP, werden als Scorers klassifiziert, die restlichen 45 Prozent als Aggressors (vgl 2.3).

Tabelle 8: Kategorisierung der Verteidiger nach Spielertypen

Spielertyp	Anteil	n	PTSPGP	PLUSMINUSPGP	PIMPGP	WEIGHT
Scorers (D)	55%	291	0.405	0.030	0.485	92.57
Aggressors	45%	238	0.149	-0.019	0.559	93.53
<i>Total</i>	<i>100%</i>	<i>529</i>	<i>0.290</i>	<i>0.011</i>	<i>0.514</i>	<i>93.00</i>

Die Aufteilung der Verteidiger in die zwei Kategorien Scorers, und Aggressors ergibt Durchschnittswerte, die mit den Ergebnissen von Vincent und Eastman (2009a) vergleichbar sind. PTSPGP fällt im Mittel für diesen Datensatz für Scorers 0.055 höher aus. Für Aggressors sind die Werte identisch. Die Variable PLUSMINUSPGP ist für Scorers ebenfalls mit den Werten von Vincent und Eastman (2009a) vergleichbar. Aggressors hingegen weisen in diesem Datensatz eine deutlich bessere Plus/Minus-Bilanz auf (+0.591). PIMPGP haben sich über alle Verteidiger hinweg gesehen halbiert, wobei die Veränderung der Aggressors deutlich höher ausfällt als bei den Scorers (-0.185 für Scorers und -0.871 für Aggressors). Früher waren Aggressors deutlich schwerer als Scorers, in den zwei beobachteten Saisons war die Variable WEIGHT allerdings nur marginal unterschiedlich. Der beobachtete Trend bei Stürmern von tieferen PIMPGP, sowie die Konvergenz des Körpergewichts trifft demnach auch auf Verteidiger zu.

Die Durchschnitte der Leistungskennzahlen für die beiden Verteidigerkategorien weichen mehr von jenen von Vincent und Eastman (2009a, S. 16) ab, als jene für Stürmer. Die Klassifizierung nach Scorers und Aggressors ergibt dennoch zwei heterogene Gruppen, welche als Regressionsgrundlage dienen können.

5.3.3 Regressionen

Die Regressionsresultate in Tabelle 9 basieren auf den Kategorisierungen, die aus den Kapitel 5.3.1 und 5.3.2 hervorgehen. Die Regressionen sollen prüfen, ob eine Einteilung nach Spielertypen zu homogeneren Kategorien und somit besseren Regressionsresultaten führt, als eine Klassifizierung nach Positionen. Die Regressionsgüte R^2 und die Anzahl signifikanter Variablen gibt darüber Aufschluss. Untenstehende Tabelle zeigt für jede Regression die Anzahl Beobachtungen n , die Regressionsgüte R^2 und die p -Werte für jede Leistungskennzahl und jede Kategorie. Variablen mit einem bedeutsamen Einfluss auf das Salär sind blau eingefärbt. Weiterführende Informationen aus den Regressionen können dem Anhang entnommen werden. Die Diskussion der Regressionsresultate erfolgt in Kapitel 5.4.

Tabelle 9: p-Werte der Regressionen nach Spielertypen

Variablen	Scorers	Grinders	Enforcers	Scorers (D)	Aggressors
<i>n</i>	316	580	122	291	238
<i>R</i> ²	0.475	0.463	0.241	0.591	0.433
AGE	0.000	0.000	0.000	0.000	0.000
GP	0.000	0.000	0.818	0.261	0.004
DRAFT	0.003	0.000	0.375	0.000	0.001
PTS	0.000	0.000	0.846	0.000	0.878
SOP	0.109	0.102	0.009	0.032	0.131
PLUSMINUS	0.126	0.001	0.952	0.038	0.207
PIM	0.272	0.523	0.433	0.366	0.690
WEIGHT	0.072	0.173	0.465	0.238	0.361
HES	0.009	0.298	0.330	0.807	0.402
SP	0.009	0.012	0.996	0.221	0.510

5.4 Interpretation & Diskussion der Resultate

Tabelle 10 fasst die Erkenntnisse aus den zehn Regressionen zusammen, indem sie eine Übersicht über alle als signifikant identifizierten Variablen gibt. Der obere Teil der Tabelle zeigt die Variablen der Regressionen nach Positionen aus Kapitel 5.2 auf und der Mittelabschnitt jene der Regressionen nach Spielertypen aus Abschnitt 5.3.3. Die bedeutsamen Leistungskennzahlen der Torhüter befinden sich aus Gründen der besseren Übersicht am Ende der Tabelle. Allgemein ist zu erkennen, dass Regressionen mit einer hohen Anzahl an Beobachtungen (vgl. Tabelle 6 & 9) mehr signifikante Variablen aufweisen und somit eine höhere Aussagekraft haben, als jene mit weniger Observationen. Es ist daher anzunehmen, der Einfluss von Leistungskennzahlen auf Spielerlöhne besser bestimmt werden kann, wenn mehrere Saisons analysiert werden.

Tabelle 10: Übersicht der signifikanten Variablen für alle Regressionen

Regression	Signifikante Variablen								
C:	AGE,	GP,	DRAFT,	PTS,		PLUSMINUS,	WEIGHT,		
RW:	AGE,	GP,	DRAFT,	PTS,				HES,	
LW:	AGE,	GP,	DRAFT,	PTS,	SOP,	PLUSMINUS,	WEIGHT,		SP
D:	AGE,	GP,	DRAFT,	PTS,		PLUSMINUS,	WEIGHT,		
Scorers:	AGE,	GP,	DRAFT,	PTS,			WEIGHT,	HES,	SP
Grinders:	AGE,	GP,	DRAFT,	PTS,		PLUSMINUS,			SP
Enforcers:	AGE,				SOP,				

Scorers (D):	AGE,	DRAFT,	PTS,	SOP,	PLUSMINUS,	
Aggressors:	AGE,	GP,	DRAFT,			
GK:	AGE,	GP,	DRAFT,			SO

Die Regressionen nach Positionen zeigen, dass zu einem grossen Teil für Stürmer (C, RW, LW) und für Verteidiger (D) ähnliche Variablen einen signifikanten Einfluss auf den Lohn haben. AGE, GP, DRAFT, PTS, PLUSMINUS und WEIGHT sind über alle Positionen hinweg signifikant in Bezug auf die Höhe des Salärs. Abweichungen gibt es bei SOP. Die Spielweise wurde nur für die Position des linken Flügels als signifikantes Attribut klassifiziert. Dies könnte mit dem niedrigen Prozentsatz der Rechtsausleger auf dieser Position (13.54%, vgl. Tabelle 5) zusammenhängen. Auch die Leistungskennzahl SP ist nur für linke Flügelstürmer signifikant. Im Kontrast dazu steht die Variable HES, die nur für die Position RW einen signifikanten Einfluss auf den Lohn hat.

Die Regressionen nach Spielertypen signalisieren, dass je nach Kategorie andere Variablen einen signifikanten Einfluss auf die Höhe des Gehalts haben. Für Spieler, bei welchen die Physis des Eishockeysports im Vordergrund steht (Enforcers und Aggressors) ist die Punkteausbeute nicht signifikant, für spielerisch versierte Kategorien (Scorers, Grinders, Scorers (D)) hingegen schon. Für die beiden Stürmerkategorien mit Fokus auf das spielerische Element (Scorers, Grinders) ist zudem die Schussquote signifikant. In Einklang mit den Regressionen nach Positionen steht der signifikante Einfluss des Alters auf den Lohn. Überraschen mag weiter der signifikante Einfluss von HES für Scorers sowie auch die Signifikanz der Variable SOP für Enforcers und ihre Insignifikanz für Scorers und Grinders. Über den signifikanten Einfluss von PLUSMINUS bei Grinders und Scorers (D) gibt es ebenso keine intuitive Erklärung.

Ein Vergleich der Anzahl signifikanter Variablen zwischen den Regressionen für Spielerpositionen und Spielertypen zeigt, dass die Einteilung nach Positionen pro Regression mehr bedeutsame Regressoren, bezogen auf den Lohn, ergibt. Somit führt eine Einteilung nach Positionen zu homogeneren Kategorien als eine Gruppierung nach Spielertypen, was sich auch in der Höhe der Bestimmtheitsmasse der Regressionen (vgl. Tabelle 6 & 9) niederschlägt.

Für Torhüter haben sich mit AGE, GP und DRAFT dieselben Variablen als signifikant erwiesen wie für Feldspieler. Entgegen der Erwartungen ist die Fangquote (SPE) nicht als bedeutend für die Lohnbildung zu klassifizieren, die Anzahl Spiele ohne Gegentore

(SO) hingegen schon. Dies überrascht insofern, als dass das primäre Ziel – nämlich ein Spiel zu gewinnen – nicht davon abhängt, ob die Mannschaft ohne Gegentor bleibt, sondern viel mehr, dass sie mehr Tore erzielen kann, als sie hinnehmen muss.

Die vergleichsweise geringe Anzahl an signifikanten Variablen für die Regressionen der Enforcers und Torhüter könnte auf die kleine Anzahl der Observationen zurückzuführen sein. Für Enforcers wurden 122 Datensätze analysiert, für Torhüter 134. Bemerkenswert ist weiter, dass die Variable PIM für keine der Regressionen als signifikant klassifiziert wurde, obwohl in Abbildung 9 eine lineare Abhängigkeit zwischen Salär und Anzahl Strafminuten zu erkennen ist.

Tabelle 11: Signifikante Variablen und ihre Einflussstärke auf den Lohn

Variable	C	RW	LW	D	GK	S ¹	GR ²	EF ³	SD ⁴	AG ⁵
const	-3.18	-1.30	-3.13	-3.29	-2.53	-2.33	-2.24	-1.84	-3.47	-2.95
AGE	+8.5%	+8.2%	+9.5%	+9.9%	+8.9%	+10.3%	+7.8%	+5.4%	+10.6%	+9.0%
GP	-0.5%	-0.7%	-0.6%	+0.6%	+2.0%	-1.3%	-0.9%	-	-	+1.2%
DRAFT	-7.0%	-4.2%	-2.9%	-5.0%	-4.3%	-5.4%	-4.0%	-	-5.7%	-4.2%
PTS	+2.4%	+2.7%	+2.9%	+2.2%	-	+2.2%	+3.9%	-	+2.3%	-
SOP	-	-	-23.3%	-	-	-	-	-23.6%	+15.6%	-
+/- ¹¹	-0.8%	-	-0.8%	-0.5%	-	-	-0.9%	-	-0.5%	-
PIM	-	-	-	-	-	-	-	-	-	-
WEIGHT	+1.4%	-	+1.1%	+0.7%	-	+1.0%	-	-	-	-
HES	-	+0.2%	-	-	-	+0.3%	-	-	-	-
SP	-	-	-1.5%	-	-	-3.1%	-1.4%	-	-	-
SO	-	-	-	-	-6.7%	-	-	-	-	-

¹: S = Scorers ²: GR = Grinders ³: EF = Enforcers ⁴: SD = Scorers (D) ⁵: AG = Aggressors

Tabelle 11 zeigt den prozentualen Einfluss der signifikanten Variablen (vgl. Tabelle 10) auf das Salär für jede Position und jeden Spielertyp, wenn alle anderen Variablen konstant gehalten werden. Leere Zellen bedeuten, dass die Variable für die entsprechende Position oder Spielertyp als nicht entscheidend klassifiziert wurde. Sämtliche Werte sind auf eine Dezimalstelle gerundet. Ungerundete Werte der Koeffizienten sind den Regressionsresultaten im Anhang zu entnehmen.

¹¹ Variable PLUSMINUS (Plus/Minus-Statistik)

AGE hat für sämtliche Positionen und Spielertypen einen sehr grossen, positiven Effekt auf das Gehalt. Verteidiger erhalten mit zunehmendem Alter die grössten Lohnerhöhungen (9.9%), für rechte Flügel fallen sie mit 8.2% am geringsten aus. Auf Spielertypen bezogen unterstreichen die Erhöhungen pro Altersjahr die Wichtigkeit der Scorers (+10.3% für S bzw. +10.6% für SD) für ein Team.

Dass die signifikante Variable GP nur bei drei von zehn Regressionen einen positiven Effekt hat, steht im Kontrast zu den Erkenntnissen aus der linearen Einfachregressionen in Abbildung 5. Jedoch ist die Grösse der negativen Einflüsse auf das Salär relativ gering (-0.5% bis -1.3%). Der positive Effekt einer steigenden Anzahl absolvierter Partien für Torhüter (+2.0%) könnte auf den tiefen Durchschnitt (39.05, vgl. Tabelle 5) zurückzuführen sein, der deutlich unter den arithmetischen Mittelwerten für Feldspieler liegt.

Eine schlechtere Platzierung im Entry-Draft der NHL hat die erwartete negative Auswirkung auf den Lohn. Die Einflüsse der Variable DRAFT variieren allerdings stark. Während die Löhne von linken Flügel 2.9% pro Drafrunde, in welcher sie nicht gezogen werden, tiefer ausfallen, beträgt der Abschlag für Center 7.0% pro Runde. Der Unterschied könnte damit erklärt werden, dass es in der NHL deutlich mehr Center als linke Flügelstürmer gibt, wie die Stichprobengrössen aus Tabelle 5 beweisen. Weiter ist zu erwähnen, dass die Variable DRAFT – ungleich vieler anderer signifikanter Variablen – vom Spieler selber kaum beeinflussbar ist. Im Gegensatz zu GP, PTS, PLUSMINUS, PIM oder SP kann sie auch nicht jede Saison zugunsten eines höheren Salärs beeinflusst werden, da jeder Spieler nur einmal, nämlich am Anfang seiner Karriere, gedraftet wird.

Eine höhere Punkteausbeute (PTS) in der Regular Season führt zu einem höheren Salär. Dieses Resultat war insofern zu erwarten, als dass Statistiken über Skorerpunkte sehr oft als erster Schätzer für die Fähigkeiten eines Spielers herangezogen werden. Auch der höhere positive Einfluss für die Stürmerpositionen C, RW und LW gegenüber Verteidiger scheint logisch, da Stürmer in erster Linie für das Erzielen von Toren bezahlt werden. Als überraschend kann der grösste positive Einfluss für Grinders (+3.9%) bezeichnet werden. Linke Flügelstürmer und Enforcers verdienen gemäss den multiplen Regressionsresultaten rund 23 Prozent weniger, wenn sie Rechtsausleger sind. Konträr dazu verdienen als Scorers kategorisierte Verteidiger 15.6 Prozent mehr, wenn sie nach rechts ausgelegt spielen. Diese Resultate sind überraschend, insbesondere deswegen, weil rechtsauslegende Spieler mit insgesamt 38.53 Prozent Anteil deutlich rarer sind als Linksausleger.

Die multiple Regression steht im Einklang mit der Statistik über linke Flügelstürmer. Nur gerade 13.54 Prozent spielen auf dieser Position nach rechts ausgelegt. Wie die erheblichen Lohneinflüsse für Enforcers und verteidigende Scorers erklärt werden könnte, ist nicht Gegenstand dieser Arbeit.

Eine gute Plus/Minus-Bilanz sollte gemäss dem sportökonomischen Sachverstand zu einem besseren Salär führen, da der Spieler dem Team insgesamt mehr Tore einbringt, als er zulässt. Die multiple Regression zeigt jedoch das Gegenteil. Unterstützt wird dieses Resultat durch die Abbildung 8, welche die Ambiguität dieser Variable aufzeigt. Der negative Einfluss fällt mit maximal -0.9% für Grinders allerdings äusserst klein aus.

Die Resultate der multiplen Regressionen widersprechen betreffend der Variable PIM zu einem gewissen Masse den linearen Einfachregressionen aus Abbildung 9. Diese zeigen eine positive Korrelation zwischen der Anzahl Strafminuten und dem Salär. In den multiplen Regressionen wurde PIM jedoch nie als signifikante Variable klassifiziert.

Für Mittelstürmer, linke Flügel, Verteidiger und positionsübergreifend für angreifende Scorers wirkt sich ein höheres Gewicht positiv auf das Gehalt aus. Der Koeffizient der Variable WEIGHT ist für Center mit 1.4 Prozent am höchsten. Bei Verteidigern ergibt ein zusätzliches Kilogramm an Körpergewicht 0.7 Prozent mehr Lohn.

Die Anzahl Checks, die ein Spieler ausführt, sagt viel über seine Spielweise aus. Es erscheint somit logisch, dass die Variable HES insbesondere für Enforcers oder Aggressors als signifikant gelten würde. Die multiplen Regressionen für diese beiden Kategorien stuften diese Leistungskennzahl allerdings als nicht signifikant ein. Konträr dazu wirken sich mehr Checks mit 0.2 Prozent, respektive 0.3 Prozent, leicht positiv auf das Salär für rechte Flügel und scorende Angreifer aus.

Überraschend scheint der negative Einfluss einer besseren Schussquote auf das Salär. Die Variable SP wurde für linke Flügel, Scorers und Grinders als bedeutsam kategorisiert. Die negative Lohnveränderung pro Prozent besserer Schussquote – das Verhältnis von Schüssen und Toren – beträgt für linke Flügel 1.5 Prozent, für Grinders 1.4 Prozent und für Scorers 3.1 Prozent. Diese Resultate sind allerdings mit Vorsicht zu geniessen, da diese Verhältnisstatistik durch hohe Werte, die durch eine geringe Anzahl Schüsse zustandekommen, verfälscht sein kann.

Für Torhüter wurde die Variable SO als negativer Lohnfaktor identifiziert. Gemäss der multiplen Regression vermindert jeder Shutout das Salär eines Torhüters um 6.7 Prozent.

Diese Erkenntnisse stehen in einem starken Kontrast zum positiven linearen Zusammenhang, der in Abbildung 13 erkennbar ist.

6 Anwendung

In Kapitel 4 wurde die Verwendung von Regressionsmodellen zur Kontrolle erwähnt. Abschnitt 6.1 analysiert bestehende Spielerverträge mit den Regressionsmodellen von Tabelle 11 um zu kontrollieren, wie stark sich die tatsächlichen Löhne der Spieler von den Regressionsgehältern abweichen. Kapitel 6.2 erläutert Anwendungsmöglichkeiten der Forschungsergebnisse für die Praxis und würdigt die Erkenntnisse dieser Arbeit kritisch.

6.1 Analyse bestehender Verträge

Mit den Regressionsdaten Tabellen 11 können bestehende Verträge analysiert werden. Tabelle 12 und 13 stellen die Saläre von Patrick Kane und Connor McDavid den Gehältern aus der Anwendung der Regressionen gegenüber.

Tabelle 12: Anwendung der Regressionen (RW, S) auf Patrick Kane (CHI)

Patrick Kane			
ID:	417	SOP:	0
SEASON:	2017/2018	GP:	82
POS:	RW	PTS:	76
SAL:	10.5	PLUSMINUS:	-20
AGE:	29	PIM:	32
WEIGHT:	80.45	SP:	9.5
DRAFT:	1	HES:	18
<i>Regression_{RW}:</i>	$\ln(\hat{y}) = -1.3 + (0.082 \cdot 29) + (-0.007 \cdot 82) + (-0.042 \cdot 1) + (0.027 \cdot 76) + (0.002 \cdot 18)$ $\ln(\hat{y}) = 2.55$ $\hat{y} = 12.81$		
	$\ln(\hat{y}) = -2.33 + (0.103 \cdot 29) + (-0.013 \cdot 82) + (-0.054 \cdot 1) + (0.022 \cdot 76) + (0.01 \cdot 80.45) + (0.003 \cdot 18) + (-0.031 \cdot 9.5)$		
<i>Regression_S:</i>	$\ln(\hat{y}) = 1.773$ $\hat{y} = 5.89$		

Patrick Kane verdiente in der Saison 2017/2018 als rechter Flügel (RW) bei den Chicago Blackhawks (CHI) 10.5 Mio. US-Dollar. Gemäss der Regression für diese Spielerposition mit seinen Statistiken sollte sein Lohn bei 12.81 Mio. US-Dollar liegen. Kane wird wegen seinen guten Skorerstatistiken (76 Punkte) zudem als Scorer kategorisiert. Die Regression

für Scorers ergibt einen Lohn von 5.89 Mio. US-Dollar. Der Vergleich der Regressionslöhne mit dem tatsächlichen Salär zeigt auf, warum das Bestimmtheitsmass R^2 der Regressionen nur 0.570 für rechte Flügel und 0.475 für Scorers beträgt. Der Errorterm ε beträgt bei Patrick Kane -2.31 Mio. US-Dollar für die Regression der rechten Flügel, beziehungsweise 4.61 Mio. US-Dollar. für jene der Scorers. Die Regression für rechte Flügel klassifiziert Kane als unterbezahlt, jene der Scorers aber als überbezahlt.

Tabelle 13: Anwendung der Regressionen (C, S) auf Connor McDavid (EDM)

Connor McDavid			
ID:	538	SOP:	1
SEASON:	2017/2018	GP:	82
POS:	C	PTS:	108
SAL:	0.925	PLUSMINUS:	20
AGE:	21	PIM:	26
WEIGHT:	87.27	SP:	15.0
DRAFT:	1	HES:	28
<i>Regression_C:</i>	$\ln(\hat{y}) = -3.18 + (0.085 \cdot 21) + (-0.005 \cdot 82) + (-0.07 \cdot 1) + (0.024 \cdot 108) + (-0.008 \cdot 20) + (0.014 \cdot 87.27)$ $\ln(\hat{y}) = 1.7787$ $\hat{y} = 5.92$		
<i>Regression_S:</i>	$\ln(\hat{y}) = -2.33 + (0.103 \cdot 21) + (-0.013 \cdot 82) + (-0.054 \cdot 1) + (0.022 \cdot 108) + (0.01 \cdot 87.27) + (0.003 \cdot 28) + (-0.031 \cdot 15.0)$ $\ln(\hat{y}) = 1.5807$ $\hat{y} = 4.86$		

Connor McDavid erzielte in der Saison 2017/2018 den Bestwert von 108 Skorerpunkten. Wie bereits in Kapitel 5.1.2.3 erwähnt, führte er diese Statistik an. Jeder Skorerpunkt erhöht in der Regression für Mittelstürmer den Lohn um 2.4 Prozent. Für McDavid resultiert dies in einen positiven Lohneinfluss von 13.36 Mio. US-Dollar. Verglichen mit Kane wird sein Lohn aufgrund seines Alters nur begrenzt positiv beeinflusst, da er acht Jahre jünger ist. Negativ wirkt sich paradoxerweise die hohe Plus/Minus-Bilanz, sowie die Schussquote von McDavid aus. Der negative Einfluss der Position im Draft wurde aufgrund seiner Ziehung in der 1. Runde minimiert. Nach Berücksichtigung aller signifikanten Einflussfaktoren sollte sein Lohn gemäss der Regression für Center 5.92 Mio. US-Dollar betragen. Für McDavid ist die Spielertypregression der Scorer anwendbar. Die Regression ergibt einen Lohn von 4.86 Mio. US-Dollar. Dennoch verdiente der Kanadier

in der Saison 2017/2018 nur 925'000 US-Dollar. Diese erhebliche Abweichung von Regressionsgehalten kann damit erklärt werden, dass McDavid aufgrund seines jungen Alters einen Entry-Level Contract¹² besitzt.

Auch wenn ein Entry-Level Contract keine Lohndiskriminierung im eigentlichen Sinne darstellt (vgl. 2.2), so beeinträchtigt er dennoch die Aussagekraft der Regressionen, indem er den Errorterm erheblich ansteigen lässt.

6.2 Anwendungsmöglichkeiten in der Praxis & kritische Würdigung

Obwohl die in dieser Arbeit konstruierten Regressionsmodelle die Löhne der NHL-Spieler eher schlecht zu erklären scheinen (vgl. 6.1), sind die Erkenntnisse über die Einflussfaktoren in der Praxis dennoch brauchbar. Die Regressionskoeffizienten können bei der Ausarbeitung von variablen Lohnkomponenten helfen. Insbesondere die Forschungsergebnisse in Bezug auf den Einfluss von Skorerstatistiken liefern hierfür gute Ansätze.

Abschliessend sei erwähnt, dass eine Quantifizierung des Eishockeysports äusserst schwierig ist. Die Regressionsmodelle sind darum kritisch zu hinterfragen, da sie bewusst viele Faktoren ausser Acht lassen, die nicht, oder nur mit grossem Aufwand quantifizierbar sind.

Eine weitere Herausforderung stellen verletzte Spieler dar. Verletzungen führen dazu, dass ein Spieler weniger Partien absolviert und damit zwangsläufig auch weniger Skorerpunkte sammelt. Diese Umstände beeinflussen zwei Regressoren negativ, wohingegen der Lohn als Regressand unverändert bleibt. Diese Probleme könnten reduziert werden, wenn der zu analysierende Datensatz vergrössert wird. Dies würde eine zusätzliche Filterung erlauben. So könnten beispielsweise verletzte Spieler, oder Spieler mit einem Entry-Level Contract von der Analyse ausgeschlossen werden.

Die Datenanalyse hat gezeigt, dass der Eishockeysport trotz unzähligen Statistiken zu grossen Teilen unberechenbar ist. Dieser Irrationalität und Emotionalität verdankt dieser Sport jedoch auch die grosse Spannung und die Tatsache, dass grosse Ligen wie die NHL viele Leute in ihren Bann ziehen und somit – wie in einleitend erwähnt – grosse Millionenbeträge umsetzen.

¹² Spieler im Alter von unter 25 Jahren erhalten einen Entry-Level Contract als erster Vertrag in der NHL, die mit einem Gehalt von maximal 925'000 US-Dollar dotiert sein dürfen (CapFriendly, 2019).

7 Literaturverzeichnis

- Balloch, G. (2014). *GSAA: An Essential Statistic For Evaluating Goaltenders*. Abgerufen von <https://ingoalmag.com/analysis/gsaa-essential-statistic-evaluating-goaltenders>.
- CapFriendly (2019). *Frequently Asked Question*. Abgerufen von <https://www.capfriendly.com/faq>.
- Eastman, B. D. (1981). The Labor Economics of the National Hockey League. *Atlantic Economic Journal*, 9(1), S. 100-101.
- Eishockey.com (2019). *NHL-Draft*. Abgerufen von http://www.eishockey.com/nhl/?-page_id=119.
- Gramarcy, R., Jensen, S., & Taddy, M. (2012). Estimating Player Contribution in Hockey with Regularized Logistic Regression. *Journal of Quantitative Analysis in Sports*, (9)1, S. 97-111.
- Hockey Reference (2018). *2017-18 NHL Skater Statistics*. Abgerufen von https://www.hockey-reference.com/leagues/NHL_2018_skaters.html.
- Hockey Reference (2018). *2018 NHL All-Star Game Rosters*. Abgerufen von https://www.hockey-reference.com/allstar/NHL_2018_roster.html.
- Hofer, M. (2017). *How Does the Salary Cap Work*. Abgerufen von <https://swisshockey-news.ch/index.php/shb/65-nhl-explained/8124-how-does-the-salary-cap-work>.
- Idson, T. L., & Kahane, L. H. (2000). Team Effects on Compensation: An Application to Salary Determination in the National Hockey League. *Economic Inquiry*, 38(2), S. 345-357.
- InvestSnips (2017). *Sports Teams Stocks U.S. Exchanges*. Abgerufen von <http://investsnips.com/list-of-publicly-traded-sports-franchises>.
- Jones, J. C., & Walsh, W. D. (1988). Salary Determination in the National Hockey League: The Effects of Skills, Franchise Characteristics, and Discrimination. *Industrial & Labor Relations Review*, 41(4), S. 592-604.
- Jones, J. C., Nadeau, S., & Walsh, W. D. (1997). The Wages of Sin: Employment and Salary Effects of Violence in the National Hockey League. *Atlantic Economic Journal*, 25(2), S. 191-206.
- Jones, J. C., Nadeau, S., & Walsh, W. D. (1999). Ethnicity, Productivity and Salary: Player Compensation and Discrimination in the National Hockey League. *Applied Economics*, 31(5), S. 593-608.
- Kahane, L. H. (2001). Team and Player Effects on NHL Player Salaries: A Hierarchical Linear Model Approach. *Applied Economics Letters*, 8(9), S. 629-632.

- Lavoie, M., & Grenier, G. (1992). Discrimination and Salary Determination in the National Hockey League: 1977 and 1989 Compared. In: G. W. Scully (Hrsg.): *Advances in the Economics of Sport*. S. 151-175. Greenwich: Emerald Group Publishing Limited.
- Lavoie, M., Grenier, G., & Coulombe, S. (1992). Performance Differentials in the National Hockey League: Discrimination versus Style-of-Play Thesis. *Canadian Public Policy/Analyse de Politiques*, 18(4), S. 461-469.
- Leadley, J. C., & Zygmunt, Z. X. (2006). National Hockey League. In: J. Fizek (Hrsg.): *Handbook of Sports Economics Research*. S. 49-98. New York, NY: M.E. Sharpe.
- Longley, N. (1995). Salary Discrimination in the National Hockey League: The Effects of Team Location. *Canadian Public Policy/Analyse de Politiques*, 21(4), S. 413-422.
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), S. 281-302.
- Montgomery, C. D., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. 5. Auflage. Hoboken, NJ: John Wiley & Sons.
- Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for Business and Economics*. 8. Auflage. Boston, MA: Pearson.
- Ozarian, M., & Badenhausen, K. (2018). *The NHL's Most Valuable Teams*. Abgerufen von <https://www.forbes.com/sites/mikeozarian/2018/12/05/the-nhls-most-valuable-teams>.
- Spotrac (2018). *NHL Team Contracts & Payrolls*. Abgerufen von <https://www.spotrac.com/nhl>.
- Vincent, C. B., & Eastman, B. (2009a). Defining the Style of Play in the NHL: An Application of Cluster Analysis. *Journal of Quantitative Analysis in Sports*, 5(1), S. 1-21.
- Vincent, C., & Eastman, B. (2009b). Determinants of Pay in the NHL: A Quantile Regression Approach. *Journal of Sports Economics*, 10(3), S. 256-277.
- Vollman, R. (2018). *NHL 2017-18*. Abgerufen von <http://www.hockeyabstract.com/testimonials/nhl2017-18>.
- Vollman, R., & Cane, M. (2018). *Download Statistics: NHL Goalies 2017-18*. Abgerufen von <http://www.hockeyabstract.com/testimonials>.
- Walsh, W. D. (1992). The Entry Problem of Francophones in the National Hockey League: A Systemic Interpretation. *Canadian Public Policy/Analyse de Politiques*, 18(4), S. 443-460.
- Wiseman, F., & Chatterjee, S. (2010). Negotiating Salaries through Quantile Regression. *Journal of Quantitative Analysis in Sports*, 6(1), S. 1-13.

8 Anhang

#	Bezeichnung
Anhang 1:	Datensatz Feldspieler
Anhang 2:	Datensatz Torhüter
Anhang 3:	Python Code