A Look at *Draft By Numbers: Using Data and Analytics to Improve NHL Player Selection*

Michael Schuckers

St. Lawrence Univ., Canton, NY
Statistical Sports Consulting, LLC
schuckers@stlawu.edu
@SchuckersM
@EmpiricalSports

May 2, 2020

Hockey Analytics Night in Where?

NHL Player Selection

If you look at statistics and point to a column and say, 'We're drafting this guy' — have fun. I hope you're in my division.
- Brian Burke, MIT Sloan Sports Analytics Conference, 2013

Source: Article by Dave Feschuk, *The Star*, March 1, 2013

Primary Goals

- **Paper** Take historical data publicly available to teams at the time of the draft and build a model to compare with how teams did. *Can we draft better than team's with just analytics?*

Primary Goals

- **Paper** Take historical data publicly available to teams at the time of the draft and build a model to compare with how teams did. *Can we draft better than team's with just analytics?*
- **Talk** Walk through some of the methods used in the paper, particularly Generalized Additive Models (GAM's) and interactions.

Primary Goals

- **Paper** Take historical data publicly available to teams at the time of the draft and build a model to compare with how teams did. *Can we draft better than team's with just analytics?*
- **Talk** Walk through some of the methods used in the paper, particularly Generalized Additive Models (GAM's) and interactions.

Primary Goals

- **Paper** Take historical data publicly available to teams at the time of the draft and build a model to compare with how teams did. *Can we draft better than team's with just analytics?*
- **Talk** Walk through some of the methods used in the paper, particularly Generalized Additive Models (GAM's) and interactions.

Version of the Paper

NHL Entry Draft

- Annually at the end of June over two days
- Method by which newly eligible players are allocated to NHL teams
- Eligible players: 18 years old on or before September 15 and not older than 20 years old before December 31 (Worldwide)
- Each team (of 31) starts with a pick in each of 7 rounds; picks are tradeable
- Team with the worst record has best chance at first pick in first round (lottery among non-playoff teams)
- After 1st round, picks are made in order of reg. season finish

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)
7. WPG: ????

Who should WPG select? Consensus Available (in order):

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)
7. WPG: ????

Who should WPG select? Consensus Available (in order):

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)
7. WPG: ????

Who should WPG select? Consensus Available (in order):
Sean Couturier (F), Dougie Hamilton (D), Ryan Murphy (D), Sven Baertschi (F), Nathan Beaulieu (D), Duncan Siemens (D), Mark Scheifele (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)
7. WPG: **Mark Scheifele (F)**

Who should WPG select? Consensus Available (in order):
Sean Couturier (F), Dougie Hamilton (D), Ryan Murphy (D), Sven Baertschi (F), Nathan Beaulieu (D), Duncan Siemens (D), Mark Scheifele (F)

Motivation the 2011 NHL Draft

The 2011 NHL Draft, first six picks

1. EDM: Ryan Nugent-Hopkins (F)
2. COL: Gabriel Landeskog (F)
3. FLA: Jonathan Huberdeau (F)
4. NJD: Adam Larsson (D)
5. NYI: Ryan Strome (F)
6. OTT: Mika Zibanejad (F)
7. WPG: **Mark Scheifele (F)**
8. PHL: Sean Couturier (F)
9. BOS: Dougie Hamilton (D)
10. MIN: Jonas Brodin (D)
11. COL: Duncan Siemens (D)

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise
- Some analytics about Couturier $>>$ Scheifele

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise
- Some analytics about Couturier $>>$ Scheifele
- Post draft: Couturier had 77, 46 GP (Games Played) **1st 2 seasons**

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise
- Some analytics about Couturier >> Scheifele
- Post draft: Couturier had 77, 46 GP (Games Played) **1st 2 seasons**
- Post draft: Scheifele had 7, 4 GP **1st 2 seasons**

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise
- Some analytics about Couturier >> Scheifele
- Post draft: Couturier had 77, 46 GP (Games Played) **1st 2 seasons**
- Post draft: Scheifele had 7, 4 GP **1st 2 seasons**
- Today after 8+ seasons: Couturier has 647 GP, 402 Pts, 156 Goals

Motivation the 2011 NHL Draft

- Consensus that Scheifele pick was a moderate surprise
- Some analytics about Couturier >> Scheifele
- Post draft: Couturier had 77, 46 GP (Games Played) **1st 2 seasons**
- Post draft: Scheifele had 7, 4 GP **1st 2 seasons**
- Today after 8+ seasons: Couturier has 647 GP, 402 Pts, 156 Goals
- Today after 8+ seasons: Scheifele has 519 GP, 444 Pts, 180 Goals

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, **noisy and sparse data** with slow feedback and

Motivation the 2011 NHL Draft

Illustrate some themes:

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, **noisy and sparse data** with slow feedback and
- (to teams) results incredibly important

NHL Player Selection

Desjardins (2004-), NHL League Equivalencies (NHLe),

- Estimate $\hat{\rho}_\ell$ where $\hat{Y}_{NHL,t} = \hat{\rho}_\ell Y_{\ell,t-1}$
- ratio estimator
- for each league, ℓ, e.g. SEL, OHL, QMJHL, NCAA
- where Y is Points generally
- t is year, $t-1$ previous year
- can be adjusted further by TOI
- allows a measure of league quality
- need amount of data

Extended by Rob Vollman, added by age $\hat{\rho}_{\ell,age}$

NHLe from 2014-15

Updated Translation Factors, as of 2014-15 NHL season
.80 Kontinental Hockey League (up .02)
.60 Swedish Hockey League (up .05)
.47 American Hockey League (up .02)
.44 Western Collegiate Hockey Association (defunct, up .02)
.41 National Collegiate Hockey Conference (new league)
.40 Switzerland NLA (up .04)
.37 Hockey-East (up .04)
.35 Big-10 (new league)
.32 Ontario Hockey League (up .02)
.32 Central Collegiate Hockey Association (defunct) (up .02)
.29 Finland SM-Liiga (down .01)
.27 Western Hockey League (up .01)
.26 Quebec Major Junior Hockey League (no change)
.23 Eastern College Athletic Conference (down .02)

source:

`http://www.hockeyabstract.com/thoughts/updatedtranslationfactors`

NHL Player Selection

Lawrence and Weissbock (2015), Prospect Cohort Success

- Nearest neighbors approach
- Inputs: League, Age, Points, Height
- Generated comparable players (cohort)
- Pct Success ($>$ 200 GP) among cohort
- Out of sample testing outperformed teams

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)
- Demographics (eg Height and Weight)

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)
- Demographics (eg Height and Weight)
- Pre-Draft Performance (eg PPG, SVPct)

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)
- Demographics (eg Height and Weight)
- Pre-Draft Performance (eg PPG, SVPct)
- Scouting (via Central Scouting Service)

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)
- Demographics (eg Height and Weight)
- Pre-Draft Performance (eg PPG, SVPct)
- Scouting (via Central Scouting Service)

Sources of Information

- Response/Outcome metrics (eg TOI in NHL)
- Demographics (eg Height and Weight)
- Pre-Draft Performance (eg PPG, SVPct)
- Scouting (via Central Scouting Service)

Evaluation of draft picks takes time esp. in Hockey

Response metrics

Goals:

- Available for all players in corpus

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions
- Length of time? Career? First 5 Years?

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions
- Length of time? Career? First 5 Years?

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions
- Length of time? Career? First 5 Years?

Choices:

- Time on Ice (TOI)

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions
- Length of time? Career? First 5 Years?

Choices:

- Time on Ice (TOI)
- Games Played (GP)

Response metrics

Goals:

- Available for all players in corpus
- Relevant to performance
- Comparable across positions
- Length of time? Career? First 5 Years?

Choices:

- Time on Ice (TOI)
- Games Played (GP)
- Cumulative First Seven Years per CBA (Schuckers and Argeris, 2015)

Demographics

Choices:

- Height
- Weight
- Position (C, F, D, G)

and functions of these.
As much as possible taken from draft eligible window.
In this NHL.com is pretty good because they don't update their site.

Pre-Draft Performance

Needs:

- Data has to be available for players in corpus
- Available for nearly every league in draft -1 years
- Need to know league drafted from

Choices:

- PPG
- GAA
- GP
- Leagues(Liiga, NCAA, OHL, *Other*, QMJHL, Russia, Russia2, USHL, WHL)

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$
  - EU Skaters $= 6.27$, EU Goalies $= 38.18$

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$
  - EU Skaters $= 6.27$, EU Goalies $= 38.18$
- To get Cescin multiply values from above by CSS Rank

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$
  - EU Skaters $= 6.27$, EU Goalies $= 38.18$
- To get Cescin multiply values from above by CSS Rank
- J Quinn 2020 #9 Ranked NA Skater would be CESCIN $= 9 \times 1.35 = 12.15$

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$
  - EU Skaters $= 6.27$, EU Goalies $= 38.18$
- To get Cescin multiply values from above by CSS Rank
- J Quinn 2020 #9 Ranked NA Skater would be CESCIN $= 9 \times 1.35 = 12.15$

Scouting Data

Via Central Scouting Player Rankings

- NHL Internal Rankings from Central Scouting Service
- Four Groups: (Skaters, Goalies) $\times$ (North Amer., Europe)
- Convert to single list via Iain Fyffe's Cescin (2011)
- Central Scouting Integrator (Cescin)
  - NA Skaters $= 1.35$, NA Goalies $= 13.25$
  - EU Skaters $= 6.27$, EU Goalies $= 38.18$
- To get Cescin multiply values from above by CSS Rank
- J Quinn 2020 #9 Ranked NA Skater would be CESCIN $= 9 \times 1.35 = 12.15$

Link to Fyffe Article, Link to 2020 CSS Rankings

Join the Data

Large Data Wrangling Effort

- Combine Data by Player & Draft Year
- Some players ranked by CSS but not drafted
- Data in two cohorts
  - 1998-00 (Training), 2001-02 (Test), 2007-08 (Validate)
  - 2004-06 (Training), 2007-08 (Test)
- Spellings, accents, multiple players same name

I was inefficient at this task at best.

Robin Olssons (Born in '89 or '90)

NAME	BORN	YEAR | LATEST TEAM
Robin Olsson (D)	05/30/1989	2015 | AIK | Allsvenskan
Robin Olsson (D)	02/05/1989	2020 | Tingsryds AIF | Allsvenskan
Robin Olsson (F)	04/07/1990	2010 | IK Graip | Division 3
Robin Olsson (F)	08/03/1995	2012 | IK Oskarshamn J18 | J18 Div.1
Robin Olsson (F)	09/24/1997	2017 | Bollnäs IS | Division 3
Robin Olsson (F)	09/19/1989	2009 | Filipstads IF | Division 3
Robin Olsson (G)	1990	2009 | Varbergs HC | Division 3
Robin Olsson (F)	10/29/1992	2012 | Stenungsund HF J18 | J18 Div.2
Robin Olsson (D)	05/16/1995	2013 | Rönnängs IK J18 | J18 Div.2
Robin Olsson (F)	02/18/2003	2021 | Sollefteå HK J18 | J18 Div.1
Robin Olsson (D)	04/15/1989	2019 | Sala HK | Division 4

Model

Warning: Notation ahead

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family
- Generalized Additive Model

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family
- Generalized Additive Model
- $f_j()$'s variety of functional forms and fits (Splines, Loess, quadratic)

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family
- Generalized Additive Model
- $f_j()$'s variety of functional forms and fits (Splines, Loess, quadratic)
- More flexible relative to regular multiple linear regression

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family
- Generalized Additive Model
- $f_j()$'s variety of functional forms and fits (Splines, Loess, quadratic)
- More flexible relative to regular multiple linear regression
- Still linear on the Right Hand Side

Model

Warning: Notation ahead

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

Predict Response Y_i (either 1st 7 GP or 1st 7 TOI) for player i

- g() is link function, we will use $g() = \log()$ via Poisson family
- Generalized Additive Model
- $f_j()$'s variety of functional forms and fits (Splines, Loess, quadratic)
- More flexible relative to regular multiple linear regression
- Still linear on the Right Hand Side
- And interactions

GAM's

Relationship between GP and Cescin

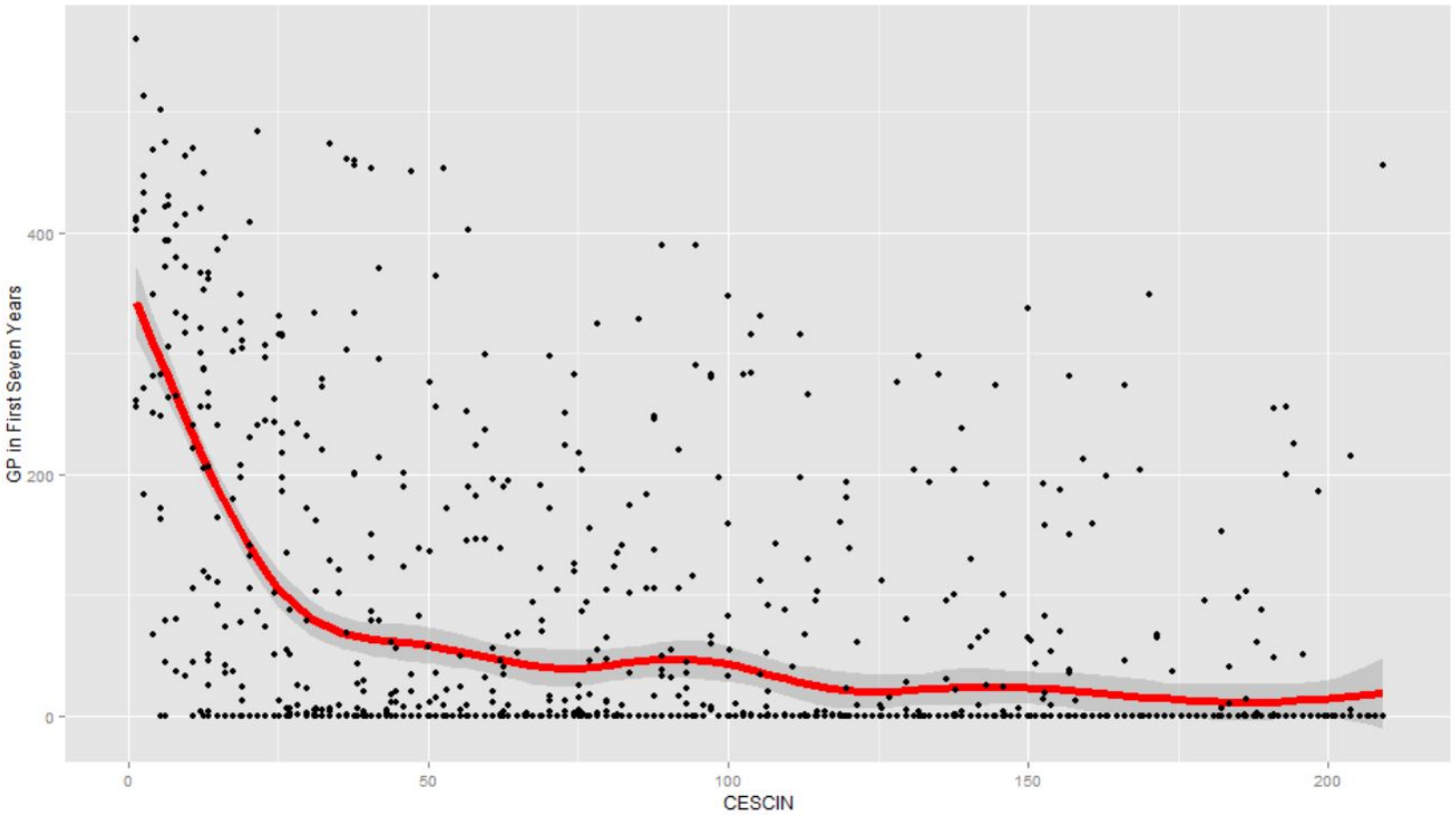

Figure 1: Plot of GP versus Cescin

GAM's

Multiple Regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

GLM

$$g^{-1}(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

GAM

$$g^{-1}(Y_i) = \beta_0 + \beta_1 f_1(X_{1i}) + \beta_2 f_2(X_{2i}) + \ldots + \beta_k f_k(X_{ki})$$

NB: There are other differences in estimation for non-Gaussian families

Additivity

Typical Linear Multiple Regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

Additivity means each effect adds to the others.
Transformation (via $g()$ or $g^{-1}()$)such as log or logistic means that the impacts are different. In particular, log transformation means a multiplicative effect.

Interactions and Indicator Variables

Interactions and Indicators allow for *nuanced* models.
Example: Indicator Variables

$$x_i = \begin{cases} 1 & \text{if player } i \text{ is a Goalie,} \\ 0 & \text{if player } i \text{ is not a Goalie} \end{cases} \tag{1}$$

Interactions are variables created by multiplying two variables together.
Eg. $X_7 = X_3 X_4$ but allows for great model flexibility

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Full (Grossly Hypothetical) Model:

$$\begin{aligned} TOI &\sim 400 + 60 I_G + 0.01 Height + 0.2 PPG \times I_F + 0.25 Height \times I_G \\ &- 0.005 \times f(Cescin) - 0.03 I_D \times I_{QMJHL} \end{aligned}$$

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Full (Grossly Hypothetical) Model:

$$\begin{aligned} TOI &\sim 400 + 60 I_G + 0.01 Height + 0.2 PPG \times I_F + 0.25 Height \times I_G \\ &- 0.005 \times f(Cescin) - 0.03 I_D \times I_{QMJHL} \end{aligned}$$

For G:

$$TOI \sim 400 + 60 + (0.01 + 0.25) Height - 0.005 f(Cescin)$$

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Full (Grossly Hypothetical) Model:

$$\begin{aligned} TOI &\sim 400 + 60 I_G + 0.01 Height + 0.2 PPG \times I_F + 0.25 Height \times I_G \\ &- 0.005 \times f(Cescin) - 0.03 I_D \times I_{QMJHL} \end{aligned}$$

For G:

$$TOI \sim 400 + 60 + (0.01 + 0.25) Height - 0.005 f(Cescin)$$

For F:

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Full (Grossly Hypothetical) Model:

$$\begin{aligned} TOI &\sim 400 + 60 I_G + 0.01 Height + 0.2 PPG \times I_F + 0.25 Height \times I_G \\ &- 0.005 \times f(Cescin) - 0.03 I_D \times I_{QMJHL} \end{aligned}$$

For G:

$$TOI \sim 400 + 60 + (0.01 + 0.25) Height - 0.005 f(Cescin)$$

For F:

$$TOI \sim 400 + 0.01 Height + 0.2 PPG - 0.005 f(Cescin)$$

Interactions: Example

I_G, I_F, I_D Indicator (**Positions**), I_{QMJHL} Indicator (**QMJHL**)

Full (Grossly Hypothetical) Model:

$$\begin{aligned} TOI &\sim 400 + 60I_G + 0.01Height + 0.2PPG \times I_F + 0.25Height \times I_G \\ &- 0.005 \times f(Cescin) - 0.03I_D \times I_{QMJHL} \end{aligned}$$

For G:
$TOI \sim 400 + 60 + (0.01 + 0.25)Height - 0.005f(Cescin)$

For F:
$TOI \sim 400 + 0.01Height + 0.2PPG - 0.005f(Cescin)$

For D:
$TOI \sim 400 + 0.01Height - 0.005f(Cescin) - 0.03I_D I_{QMJHL}$
One regression, but three different relationships

Model Evaluation

Use model to predict TOI (or GP) for each **out of sample** player in corpus. Rank order players from those.

Model Evaluation

Use model to predict TOI (or GP) for each **out of sample** player in corpus. Rank order players from those.

Calculate Spearman's Rank Correlation for:

- NHL Draft Order vs Actual TOI Order
- Predicted TOI Order vs Actual TOI Order

Results: Corpus Drafted Players

Table 4: Comparison of Rank Correlation Magnitude with NHL Performance Among Drafted Players

Training Data NHL Draft Years	Out of Sample Draft Year	NHL Performance Metric	NHL Draft Order	Draft by Numbers
1998, 1999, 2000	2001	TOI	0.366	0.603
1998, 1999, 2000	2001	GP	0.383	0.532
1998, 1999, 2000	2002	TOI	0.282	0.587
1998, 1999, 2000	2002	GP	0.348	0.536
2004, 2005, 2006	2007	TOI	0.403	0.642
2004, 2005, 2006	2007	GP	0.401	0.694
2004, 2005, 2006	2008	TOI	0.398	0.685
2004, 2005, 2006	2008	GP	0.401	0.708

Results: Corpus Drafted or CSS Ranked Players

Table 5: Comparison of Rank Correlation with NHL Performance Among All Players

Training Data NHL Draft Years	Out of Sample Draft Year	NHL Performance Metric	NHL Draft Order	Draft by Numbers Order
2004, 2005, 2006	2007	TOI	0.547	0.667
2004, 2005, 2006	2007	GP	0.547	0.670
2004, 2005, 2006	2008	TOI	0.553	0.670
2004, 2005, 2006	2008	GP	0.557	0.655
1998, 1999, 2000	**2007**	**TOI**	**0.547**	**0.650**
1998, 1999, 2000	**2007**	**GP**	**0.547**	**0.659**
1998, 1999, 2000	**2008**	**TOI**	**0.553**	**0.619**
1998, 1999, 2000	**2008**	**GP**	**0.557**	**0.616**

Prediction 2016 NHL Draft

Schuckers rank	Player	Draft spot	Team
1	Auston Matthews	1	Toronto
2	Patrik Laine	2	Winnipeg
3	**Charlie McAvoy**	**14**	**Boston**
4	Mikhail Sergachev	9	Montreal
5	Logan Brown	11	Ottawa
6	Matthew Tkachuk	6	Calgary
7	Tyson Jost	10	Colorado
8	Jakob Chychrun	16	Arizona
9	Jesse Puljujarvi	4	Edmonton
10	Adam Mascherin	38	Florida
11	Dante Fabbro	17	Nashville
12	Pierre-Luc Dubois	3	Columbus
13	Riley Tufte	25	Dallas
14	Luke Kunin	15	Minnesota
15	Max Jones	24	Anaheim

Source: Shinzawa 2018 Article

Prediction 2016 NHL Draft

16	Julien Gauthier	21	Carolina
17	Kieffer Bellows	19	NY Islanders
18	Taylor Raddysh	58	Tampa
19	Logan Stanley	18	Winnipeg
20	Alex Nylander	8	Buffalo
21	Brett Howden	27	Tampa
22	Jake Bean	13	Carolina
23	Alex DeBrincat	39	Chicago
24	Michael McLeod	12	New Jersey
25	Olli Juolevi	5	Vancouver
26	Nathan Bastian	41	New Jersey
27	Tyler Benson	32	Edmonton
28	Sam Steel	30	Anaheim
29	Boris Katchouk	44	Tampa
30	Clayton Keller	7	Arizona

Source: Shinzawa 2018 Article

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct
  - Add Combine Data, Internal Scouting Evaluation

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct
  - Add Combine Data, Internal Scouting Evaluation
  - Team quality metrics

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct
  - Add Combine Data, Internal Scouting Evaluation
  - Team quality metrics
  - Different methods CART, BART, etc.

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct
  - Add Combine Data, Internal Scouting Evaluation
  - Team quality metrics
  - Different methods CART, BART, etc.
  - Gamma link in GAM

Comments

- Corpus (which players are included) is important (Cf Table 4 & 5)
- This model is not great, lots of ways to improve
  - Add age and birthdate into model
  - Psychological metrics
  - Additional Years of Data, draft -1, draft -2, SVPct
  - Add Combine Data, Internal Scouting Evaluation
  - Team quality metrics
  - Different methods CART, BART, etc.
  - Gamma link in GAM
  - More Leagues, drop *Other* (Bayesian)

Return to Themes

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, **noisy and sparse data** with slow feedback and
- (to teams) results incredibly important

Return to Themes

- Evaluation of players is **hard**
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, **noisy and sparse data** with slow feedback and
- (to teams) results incredibly important

...and it is possible to improve drafting in the NHL via Statistics Models/Analytics.
"...[I]t is a numbers game."

CESCIN Update

Work with Stat major, Amanda Butterfield (D), St. Lawrence University '20

- Join CSS data & Draft Selection data
- Larger Data set (2003-2019)
- Tweaks to CESCIN values
- Ongoing project to build more complete data, eg. 2003

Link to Writeup
Link to Data

The End

Thanks

schuckers@stlawu.edu

Talkon 2019 Review of Analytics-based NHL Draft Work

My Other Papers in Hockey