A Puck Above the Rest:
Exploring the Effects of New Data on 2020 NHL Draft Decisions

Ashley Mullan and Lucy Ward

October 25, 2020

Advisors: S. Ventura, N. Citrone, R. Yurko

Problem Background and Objective

- The NHL draft usually runs in late June.
- COVID-19 forced the 2020 draft to run in early October.
- Multiple European leagues began prior to the draft, so 2020 prospects from these leagues had more data available.
- *Objective: Model players' future performance given their amateur performance and assess the impact the of additional data on their value*

Data

- Season-level data from amateur and professional seasons from 2010 to 2020
- Player characteristics and statistics
- Team-level statistics (games played, goals scored, goals against)
- New metrics (relative age, PTPP)

ID	Name	DOB	Country	HT	WT	Position	Shoots	GP	G	A	P	PM
199870	Jake Guentzel	1994-10-06	USA	180	82	LW/C	L	60	29	44	73	13

Data Modification and Response Metric

- Player statistics scaled by number of games played
- Response Metric: Professional Total Point Percentage (PTPP)

$$Player\ Points\ per\ Game = \frac{player's\ total\ season\ points}{player's\ total\ games\ in\ season}$$

$$Pro\ Team\ Points\ per\ Game = \frac{pro\ team's\ total\ season\ points}{games\ in\ season}$$

$$PTPP = \frac{Player\ Points\ per\ Game}{Pro\ Team\ Points\ per\ Game}$$

Prior Research

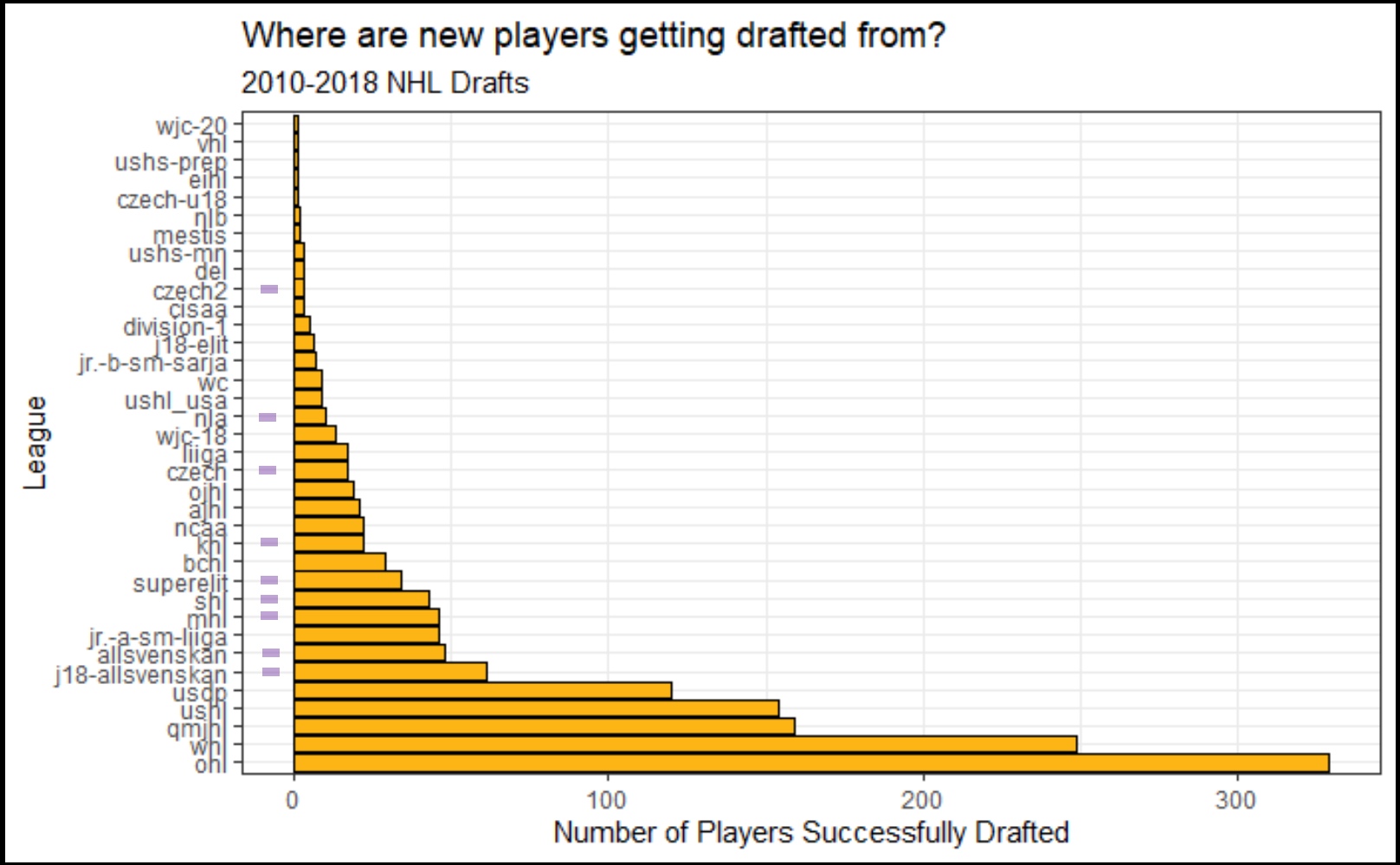
- Earlier draft choice results in better outcomes, but this effect is muted after 100 decisions. (Tingling et al, 2011)
- A Poisson GAM successfully models time on ice while including non-linear effects. (Schuckers, 2016)
- Players born in the first quarter made up the greatest percentage of the draft class. (Deaner et al, 2013)
- Early birthdays, size advantages, and anaerobic power increase chances of draft selection. (Rocznioc et al, 2013)

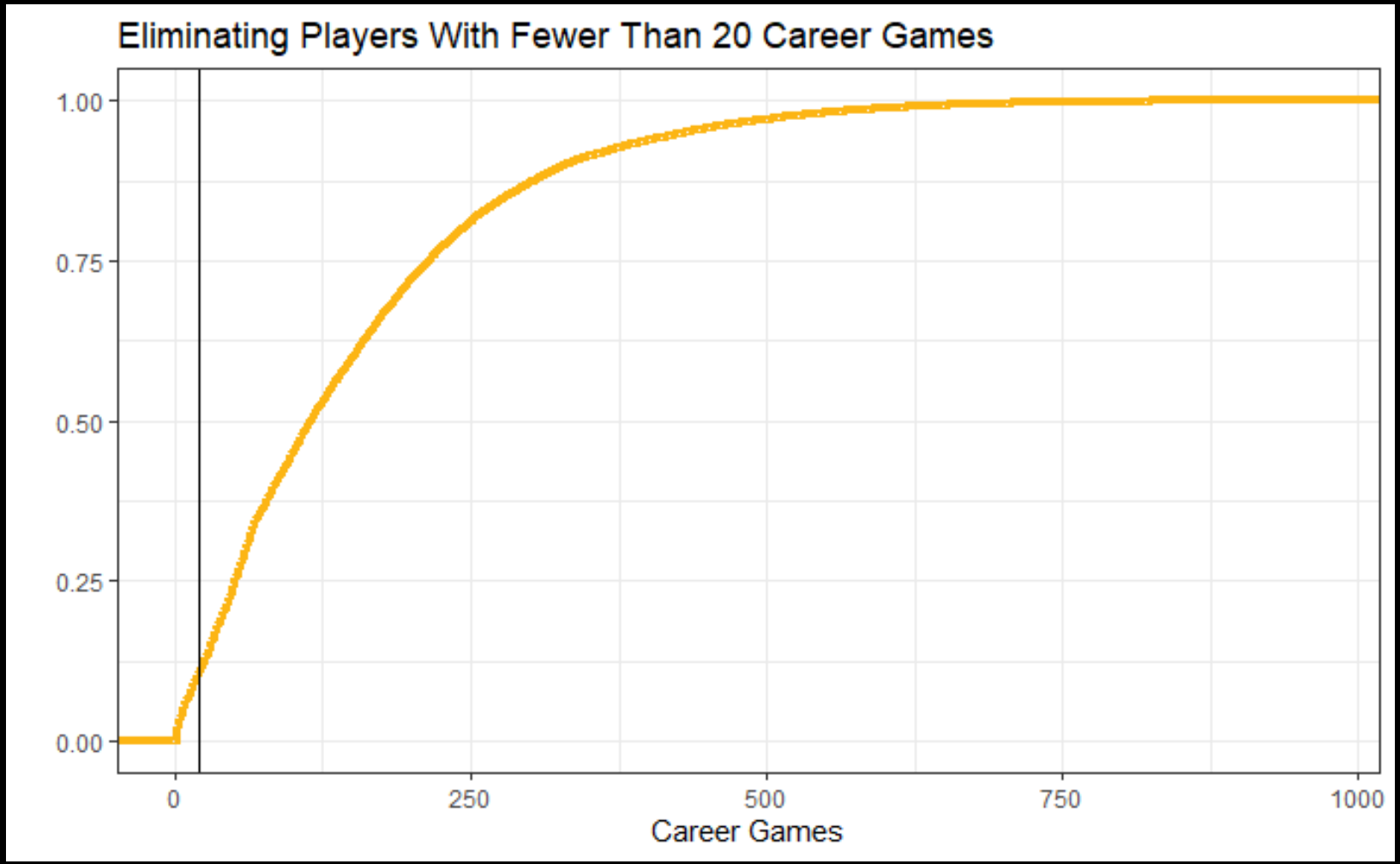
Approach

- Identify players from target leagues, both those who were drafted to the NHL and those who were not
- Model future performance given selection to NHL
- Scale expected performance by probability of making NHL
- Observe how predictions change when more data is added to simulate early Fall 2020 games

Target Leagues: Early Season Starts

- Czech Leagues
    - (Czech, Czech2)
- Russian Leagues
    - (KHL, MHL)
- Swiss Leagues
    - (NLA)
- Swedish Leagues
    - (SHL, Allsvenskan, J18-Allsvenskan, Superelit)


Where are new players getting drafted from?
2010-2018 NHL Drafts
League
wjc-20
vhl
ushs-prep
eihl
czech-u18
nlb
mestis
ushs-mn
del
czech2
cisaa
division-1
j18-elit
jr.-b-sm-sarja
wc
ushl_usa
nja
wjc-18
liiga
czech
ojhl
ajhl
ncaa
khl
bchl
superelit
shl
mhl
jr.-a-sm-liiga
allsvenskan
j18-allsvenskan
usdp
usnl
qmjhl
whl
ohl
0
100
200
300
Number of Players Successfully Drafted

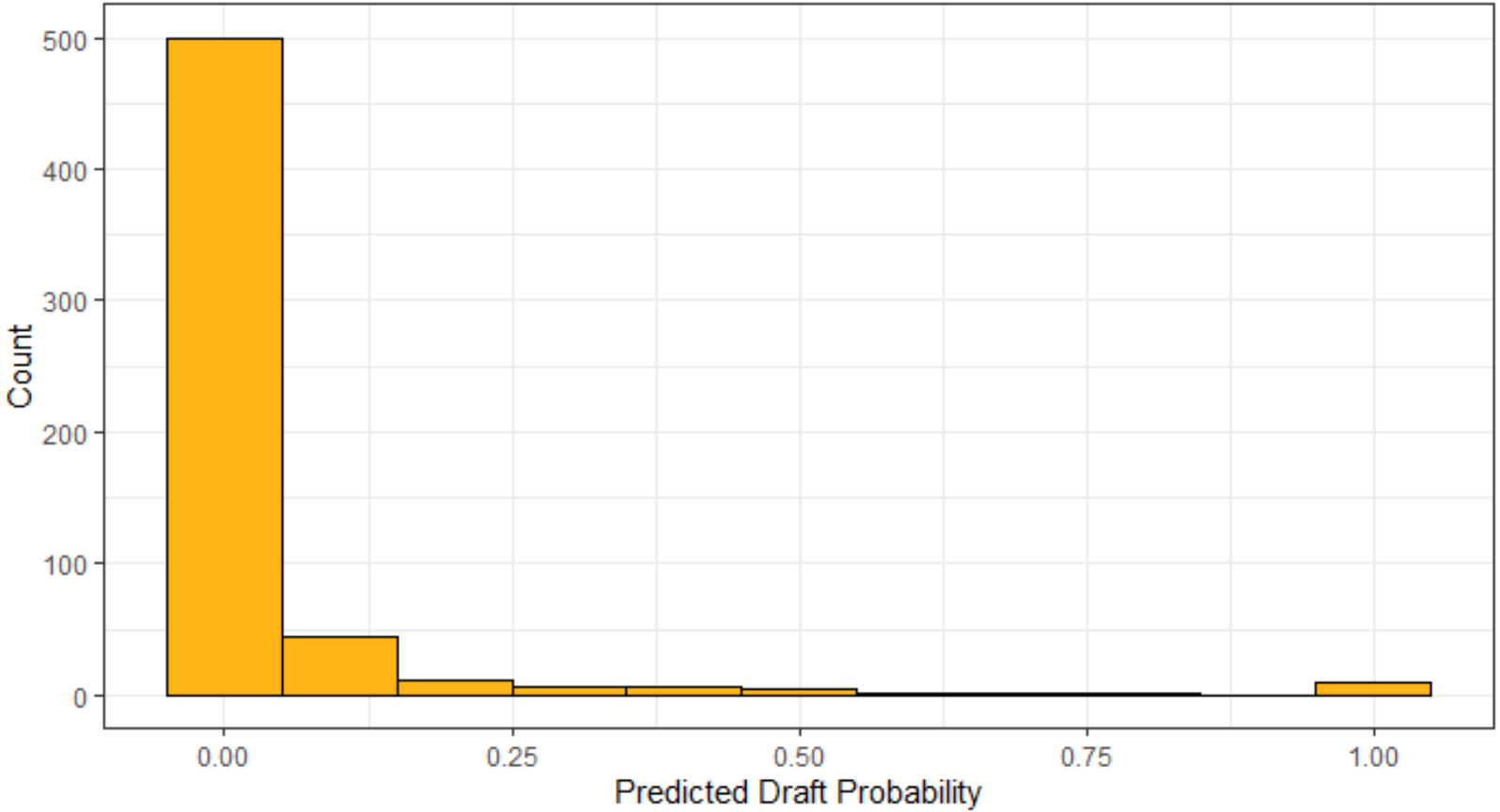

Eliminating Players With Fewer Than 20 Career Games
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
Career Games

Draft Probability Model

Goal: Predict a player's probability of being drafted into the NHL.

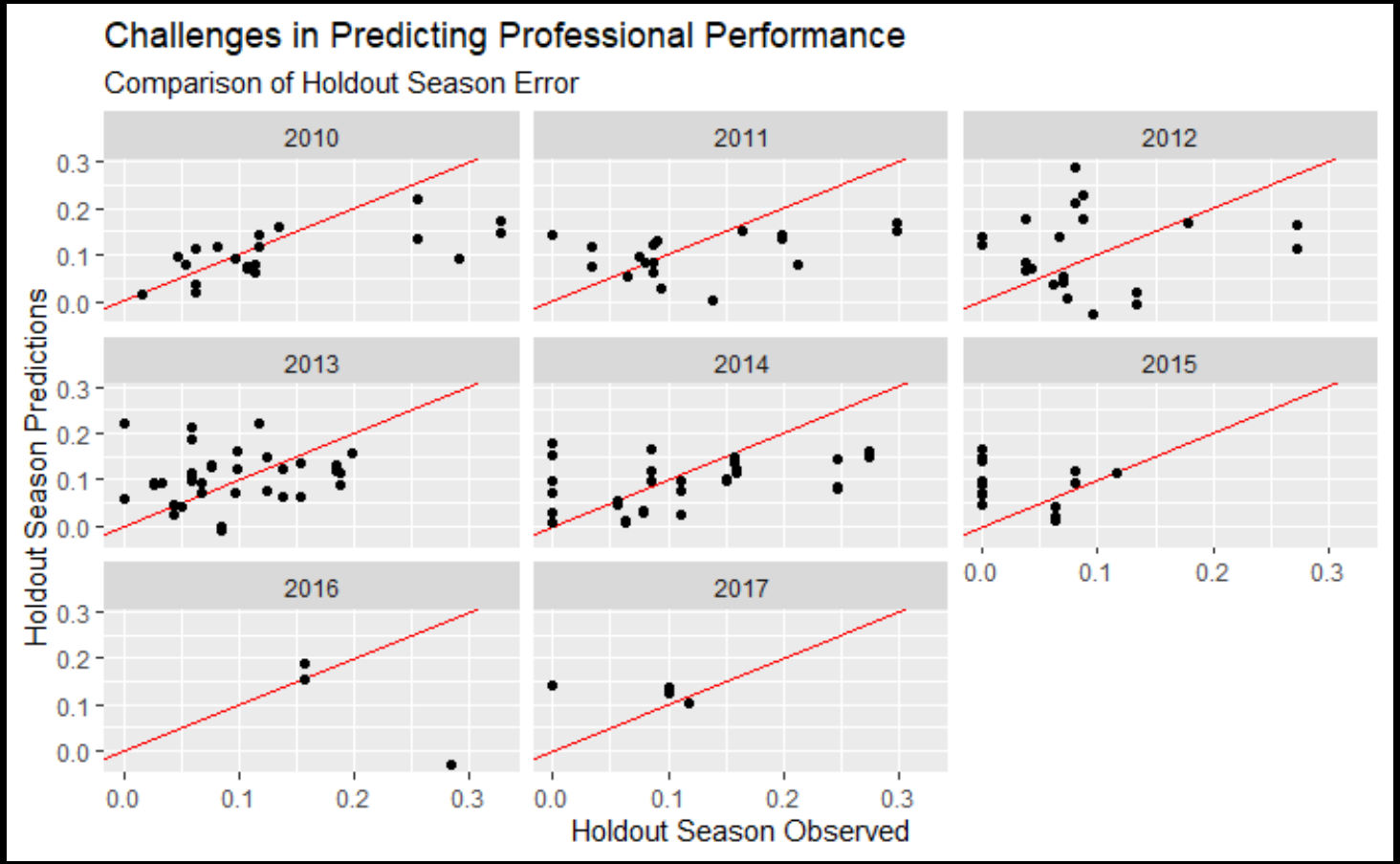
Most Players Have a Very Small Probability of Being Drafted

2018 Draft Class Prospects from Target Leagues

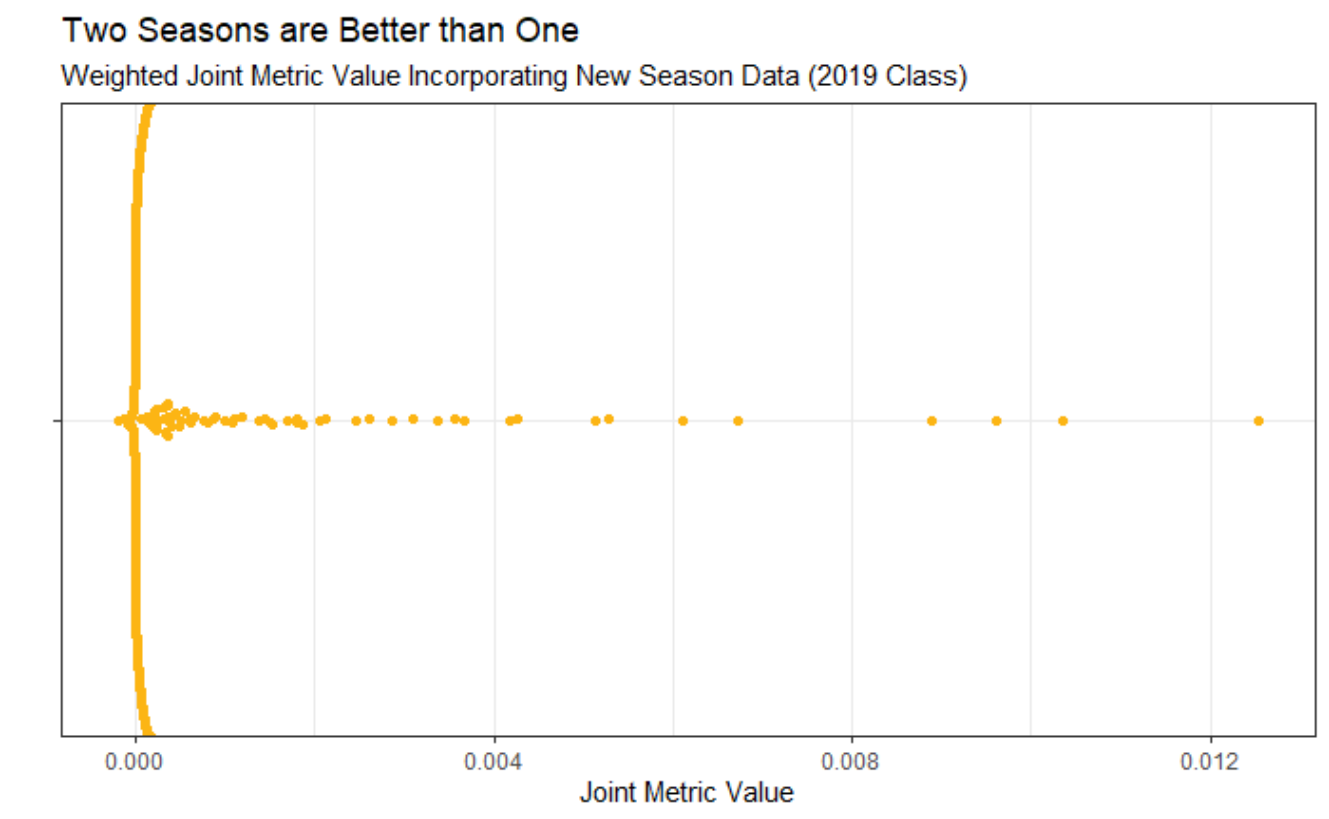


NHL Performance Model

Goal: Given that a player was drafted into the NHL, model his value to his NHL team based on his amateur statistics.


Challenges in Predicting Professional Performance
Comparison of Holdout Season Error
2010
2011
2012
2013
2014
2015
2016
2017
Holdout Season Predictions
Holdout Season Observed
0.0
0.1
0.2
0.3

Joint Metric: Combining Draft Probability with Expected Performance



Future Directions

- Improve accuracy of performance model with a potential different response variable
- Consider how to correct for lack of independence of draft probability and expected performance
- Incorporate league strength metric into player assessment to allow for broader application

Any questions?

ashley.mullan@scranton.edu
@ashley___mullan

lward7@uwyo.edu
@_lucyward_