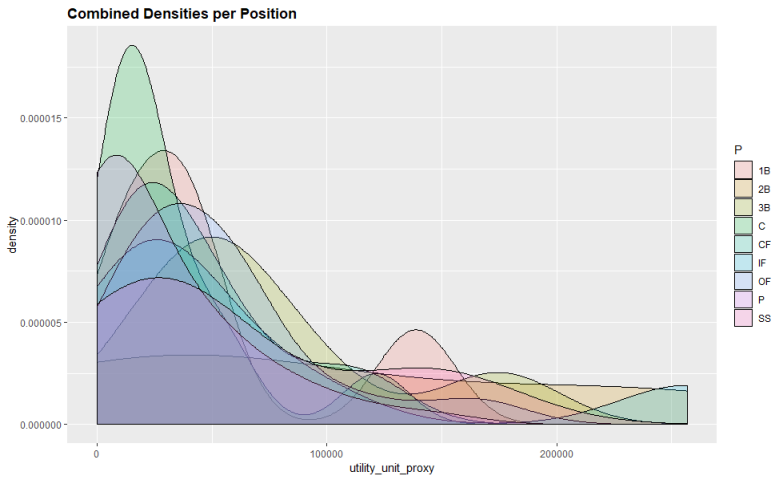
Cape Cod League Draft 2019: Model, Strategy, and Reality

Zachariah Zanger M.S.

Assistant General Manager Intern, Wareham Gatemen Baseball



Round	Name	Position	School	Dollar Amount
1	Niko Kavadas	INF	Notre Dame	65,000
2	Nick Gonzales	2B	New Mexico St.	5,000,000
3	Carmen Mlodzinski	P	South Carolina	237,000
4	Joseph Nahas	P	Georgia Southern	50,000
5	Brett Auerbach	C	Alabama	125,000
6	Zavier Warren	INF	Central Michigan	30,000
7	Jacob Teter	1B	Florida Southern	500,000
8	Noah Campbell	2B	South Carolina	100,000
9	Noah Skirrow	P	Liberty	800,000
10	Jared DeSantelo	3B	Florida Atlantic	65,000
11	Austin Love	P	North Carolina	1,000,000
12	Reid Johnston	P	NC State	500,000
13	Andrew Abbott	P	Virginia	1,125,000
14	Lucas Dunn	IF/OF	Louisville	906,800
15	Bryce Osmond	P	Jenks High School	1,000,000

Introduction

On June 22, 2019 the Wareham Gatemen Intern team drafted the Cape Cod League plus eligible Team USA players and High School players. There were 9 individuals drafting with various amounts of pool money. The draft was formatted with dollar values assigned to players eligible for the draft. In total, there were 270 players from the Cape Cod League, 59 players from High School, and 26 players from Team USA eligible for the draft. The draft consisted of 12 rounds with 3 minutes between each pick. ***For me, the draft exercise was about developing a model, using it in the draft room to guide a strategy, and experience a small taste of what the rush of a draft room is like***. In addition to daily responsibilities as Assistant GM Intern for the Wareham Gatemen, I began draft preparation two and a half weeks before the draft.

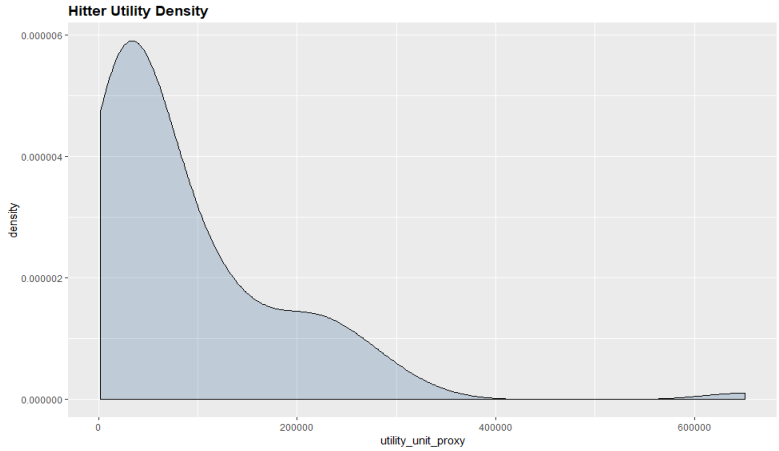
Model

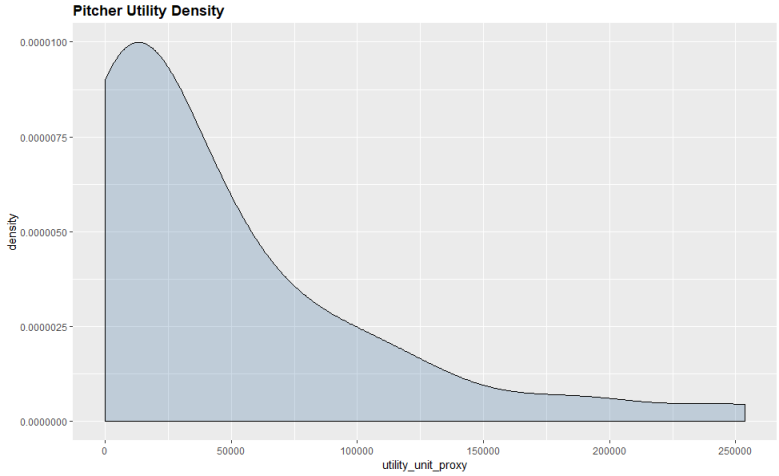
The model I developed was developed in R studio. I am more than happy to share code and supporting documents. The model draws from utility economics and marginal utilities***. I landed on developing a utility function for both hitters and pitchers, where a player's stat line comprised of certain statistics is ran through the utility function and the player earns a utility score (In the model, column name is "Utility_Unit_Proxy")***. The utility function for players by position is as follows:

$$Utility\ Function\ for\ Hitters = \left(\frac{PA}{3} + RC + (OPS * .80) + (SECA * .50) + (ISO * .50)\right)^3$$

$$Utility\ Function\ for\ Pitchers = \left(\frac{IP}{2} + K9 - (1.2 * WHIP) - ERC\right)^4$$

The only advanced stats I had for scaling purposes were the ones on the Cape Cod League website. Given the time constraints and my focus for the exercise, I did my best to create Utility functions with output that was defensible and scalable. If the soul purpose of the exercise was to produce the most intricate, best fit function for incidence of reaching the major leagues, I would have spent far more time on developing the utility function. But anyways, the distribution of utilities for both hitters and pitchers resemble power law distributions where small amounts of players stand out and there are far larger amounts of players clustered together with weaker utility scores:





The x-axis for both groups maps the quantities in utility scores. Except for the slight uptick in Hitter Utility Density around 600,000 points (more on one of those players later), these distributions have a relatively similar shape and are largely populated in the 0 – 250,000 utility score Ballpark.

I had to make sure that the scores for players were defensible. Therefore, I dialed up all the scores for hitters and pitchers, ran over the lists with scouting interns, and they seemed to agree that the lists scores were defensible. The scores for the top ten pitchers and hitters are as follows:

Pitchers

	Player	P	utility_unit_proxy	pct_change_utility
1	Seymour, I	P	253747	NA
2	Mlodzinski, C	P	247076	-0.03
3	Bedell, I	P	202187	-0.18
4	Nahas, J	P	191242	-0.05
5	Boyle, J	P	174688	-0.09
6	Dollard, T	P	151411	-0.13
7	Shuster, J	P	131643	-0.13
8	Hazelwood, M	P	126281	-0.04
9	Lardner, M	P	116000	-0.08
10	Dabovich, R	P	112323	-0.03

Hitters

	Player	P	utility_unit_proxy	pct_change_utility
1	Gonzales, N	2B	650768	NA
2	Kavadas, N	IF	318612	-0.51
3	Teter, J	1B	289220	-0.09
4	Deloach, Z	CF	278893	-0.04
5	Ward, B	CF	259902	-0.07
6	Cantrelle, H	SS	257199	-0.01
7	Schmitt, C	IF	256770	0.00
8	Cruz, T	SS	231103	-0.10
9	Baker, D	2B	230053	0.00
10	Warren, Z	IF	225755	-0.02

It should be noted that Nick Gonzales blows everyone out of the water. He is a fat tail event. Not only does the utility function believe this, but it was impossible not to see him stand out during his season for the Cotuit Kettleers. ***He lit this league up like a Christmas Tree on a snowed in Christmas.*** Nick Gonzales arguably had the best hit tool in the Cape and was arguably the best player in the league using the eye test.

The model is still not complete. There are two more themes that must be examined: Cost and the Marginal Utility added by adding one more of the player's position. With respect to Cost, the dollar amounts of players were publicized before the draft. I calculated what is essentially a utility per dollar spent figure, or as I labeled it, a Cost Score:

$$Cost\ Score = \frac{Utility\ Units}{Dollar\ Value}$$

In Finance, a Sharpe Ratio measures an expected return against the expected risk/volatility of an asset. For two assets with a similar price and expected return, the asset with the higher Sharpe ratio is deemed more attractive because of the smaller amount of risk/volatility taken on in the investment decision. With Cost Score, two players with close Utility scores could be evaluated in the context of cost/price. Examining the rate of change of utility score and price would help save money in the draft room. For instance, examining the top two pitchers in terms of utility scores (Ian Seymour and Carmen Mlodzinski)***, selecting Ian Seymour over Carmen Mlodzinski would bring a 3% higher utility score but at a price tag higher by 91%***.

	Player	P	utility_unit_proxy	Dollar.Value	cost_score	pct_change_dollars	pct_change_utility
1	Seymour, I	P	253747	2500000	0.101	NA	NA
2	Mlodzinski, C	P	247076	237000	1.043	-0.91	-0.03

The last component of the model is to attach marginal utility values to players. It is the utility value added by selecting a player in addition to a player's utility score. The marginal utility is derived from the position the player plays and the number of players at that position that have already be selected. For instance, for two players of equal utility score, similar cost, but one played first base and one played catcher, which should be selected? One way to answer is to see what is already on the roster. If there were already a catcher on the roster but no first baseman,

the 2nd catcher added may not be as valuable as the 1st first baseman. In the model, I define this value as the marginal utility score, where:

$$Marginal\ Utility\ Score = \frac{Player\ Utility\ Score}{Appropriate\ Marginal\ Utility\ Weight}$$

Marginal Utility Weights

	P	C	1B	2B	SS	INF	CF	OF
1st	1	1	1	1.25	1	1	1	1
2nd	1	1	1.25	1.5	1	1.2	1	1.2
3rd	1	1.5	2.5	2.5	1	2	1.5	1.75
4th	1	2	3	3	2	2	1.75	2
5th	2	2	3	3	3	2.5	2	2.5
6th	2	X	X	X	X	X	X	X
7th	2.5	X	X	X	X	X	X	X

These weights were not calculated with Quantitative Methods and is undoubtedly one weakness in the model (More on this soon with a potential remedy). It should be noted that in the draft list, a large portion of players were designated as simply Outfield or Infield with no specific position. I classified their position weights to be sort of like position group utilitymen. Zoning in on First Basemen Jacob Teter, Teter ranks third on the marginal benefit of selecting him:

	Player	P	Dollar.Value	utility_unit_proxy	cost_score	marginal_utility_score	pct_change_dollars	pct_change_utility
1	Gonzales, N	2B	3500000	650768	0.186	520614.4	NA	NA
2	Kavadas, N	IF	65000	318612	4.902	318612.0	-0.98	-0.51
3	Teter, J	1B	500000	289220	0.578	289220.0	6.69	-0.09
4	Deloach, Z	CF	161400	278893	1.728	278893.0	-0.68	-0.04
5	Ward, B	CF	406000	259902	0.640	259902.0	1.52	-0.07
6	Cantrelle, H	SS	165400	257199	1.555	257199.0	-0.59	-0.01
7	Schmitt, C	IF	233000	256770	1.102	256770.0	0.41	0.00
8	Seymour, I	P	2500000	253747	0.101	253747.0	9.73	-0.01
9	Mlodzinski, C	P	237000	247076	1.043	247076.0	-0.91	-0.03
10	Cruz, T	SS	400000	231103	0.578	231103.0	0.69	-0.06

But if there were a first baseman on the roster before the selection is made, Teter falls to 9th on the board:

```
> combined_cape_draft[1:10,]
```

	Player	P	Dollar.Value	utility_unit_proxy	cost_score	marginal_utility_score	pct_change_dollars	pct_change_utility
1	Gonzales, N	2B	3500000	650768	0.186	520614.4	NA	NA
2	Kavadas, N	IF	65000	318612	4.902	318612.0	-0.98	-0.51
3	Deloach, Z	CF	161400	278893	1.728	278893.0	1.48	-0.12
4	Ward, B	CF	406000	259902	0.640	259902.0	1.52	-0.07
5	Cantrelle, H	SS	165400	257199	1.555	257199.0	-0.59	-0.01
6	Schmitt, C	IF	233000	256770	1.102	256770.0	0.41	0.00
7	Seymour, I	P	2500000	253747	0.101	253747.0	9.73	-0.01
8	Mlodzinski, C	P	237000	247076	1.043	247076.0	-0.91	-0.03
9	Teter, J	1B	500000	289220	0.578	231376.0	1.11	0.17
10	Cruz, T	SS	400000	231103	0.578	231103.0	-0.20	-0.20

The marginal benefit (marginal utility score) of selecting Jacob Teter is a lot less when there is already a first baseman on the roster. In fact, the solution could be stated that the other players in front of him would carry greater marginal benefit in a selection.

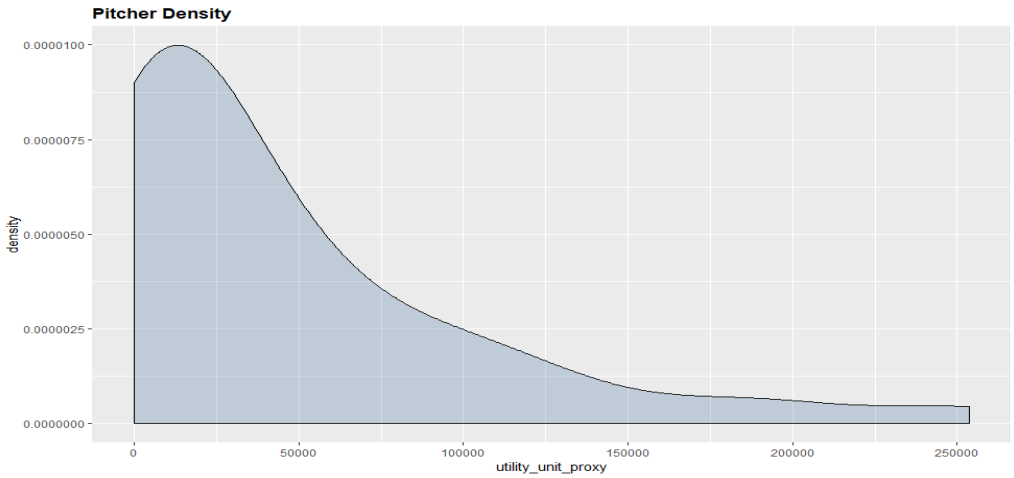
With the Marginal Utility Scores, the model is complete. The model examines a player's output from a utility function, the player's cost, and the marginal utility. The model tees up the selections. ***In respect to the position group marginal weights, if I had to enact one more***

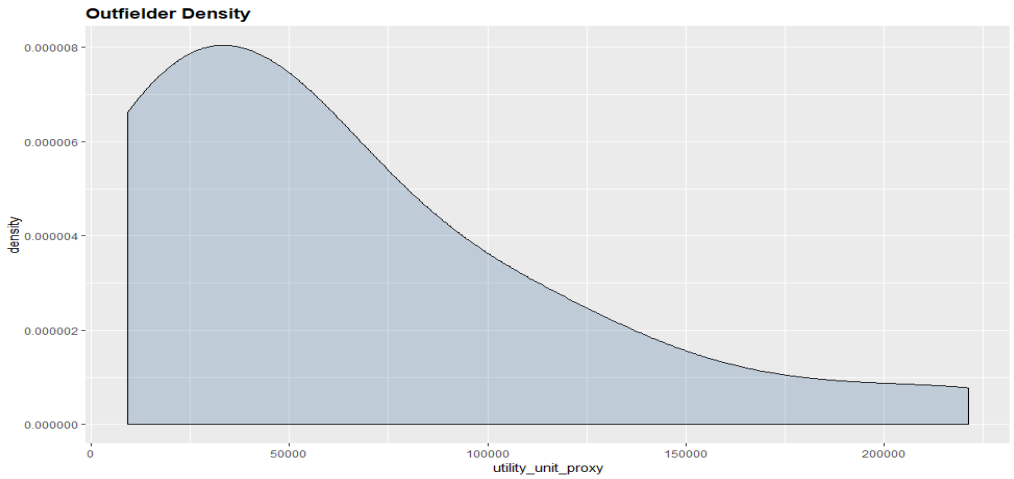
application in this project, it would be to incorporate the derivation of the marginal weights from rate of change in the quality of players in the position group. Density curves fit below, that reflect the quantity and quality of players in different groups, have differing slopes at different points along the curves. The slopes represent rate of change and could be twisted into supply functions. Supply functions with greater rates of change would represent shrinking supply for a higher quality of player. Therefore, the position groups with more rapid rate of change in decreasing quality of player should get a boost and increased in marginal utility score.

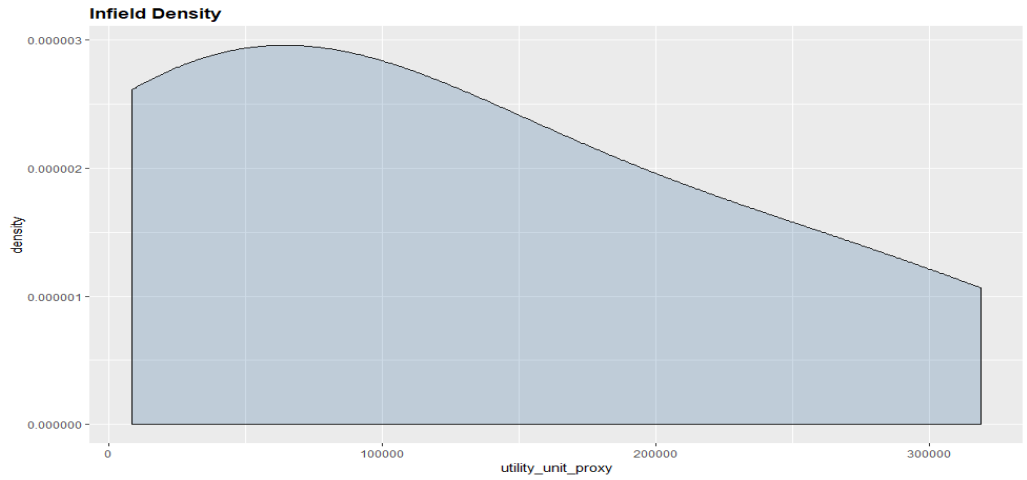
This model did not make selections. This model simply provided a path and framework for making selections. I had a strategy for making selections.

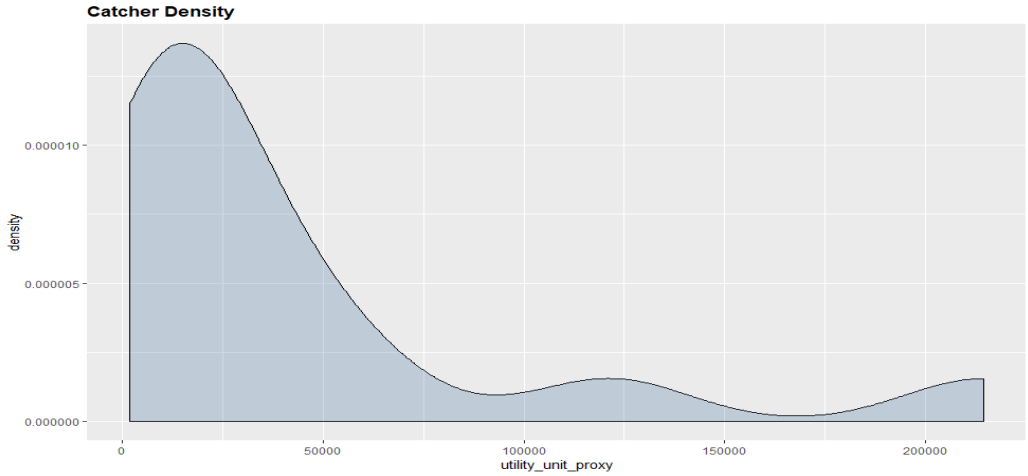
Strategy

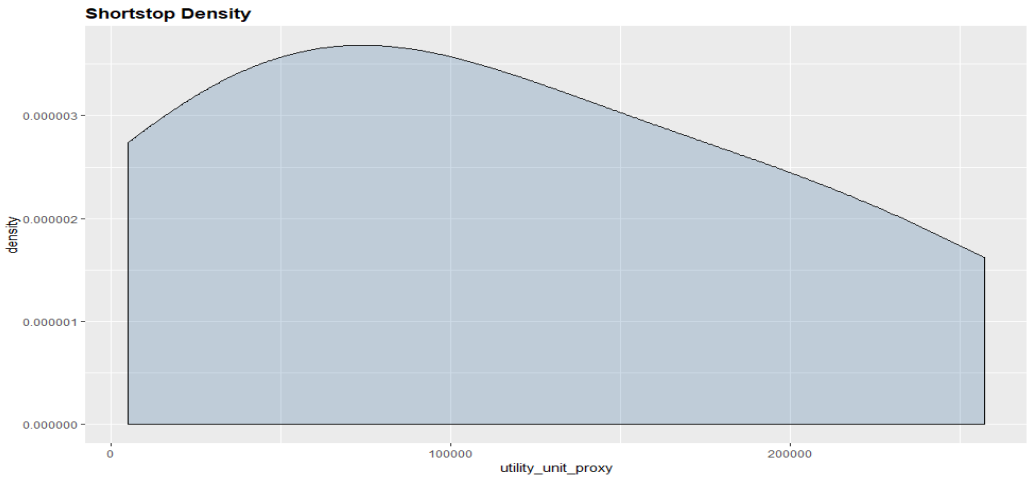
Examining the aggregate quality of player for each position group visually:

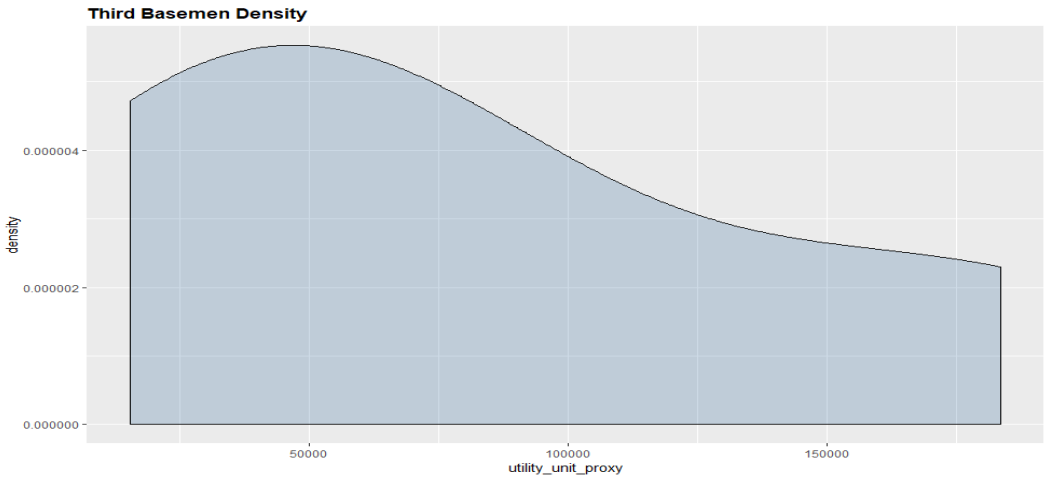


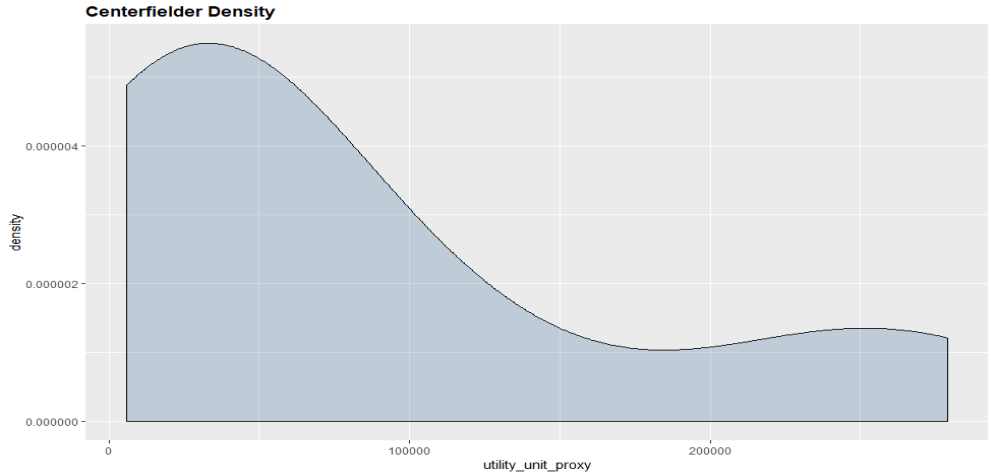



Infield Density
0.000003
0.000002
0.000001
0.000000
density
0
100000
200000
300000
utility_unit_proxy


Catcher Density
0.000010
0.000005
0.000000
density
0
50000
100000
150000
200000
utility_unit_proxy


Shortstop Density
0.000003
0.000002
0.000001
0.000000
density
0
100000
200000
utility_unit_proxy





The visuals above represent the population layouts for different utility scores (utility_unit_proxy on x-axis). A few things could be inferred from these visuals. Of the position groups, Catchers and Pitchers experienced the fastest depreciation in quality and lowest amounts of talent on the right tails of the distributions. These positions would be a priority in the early rounds.

In addition, I was confident that there were plenty of bargains throughout the draft and that the prices quoted did not reflect the intrinsic values of many players. Therefore, I would keep my eye on cost scores and change in dollar values with respect to change in utility score. I would organize my board by descending marginal utility score. I purposefully did not model in Team USA or High School players, because I needed to scale up my model consistently with the same stat lines for all players on the Cape Cod website stat page. I also believed that I would be able to draft a great roster with just Cape players while others went in other directions for players. While others were diverting their focuses among different groups of players, I was tunneling mine in to one group. I realize this is not reality, but for the purpose of building this model in a logistical sense, I made the decision to focus exclusively on Cape players.

Reality and Conclusion

```
> combined_cape_draft[1:25,]
          Player  P Dollar.Value utility_unit_proxy cost_score marginal_utility_score pct_change_dollars pct_change_utility
1    Gonzales, N 2B      3500000             650768      0.186              520614.4                 NA                 NA
2     Kavadas, N IF        65000             318612      4.902              318612.0              -0.98              -0.51
3       Teter, J 1B       500000             289220      0.578              289220.0               6.69              -0.09
4     Deloach, Z CF       161400             278893      1.728              278893.0              -0.68              -0.04
5        Ward, B CF       406000             259902      0.640              259902.0               1.52              -0.07
6   Cantrelle, H SS       165400             257199      1.555              257199.0              -0.59              -0.01
7     Schmitt, C IF       233000             256770      1.102              256770.0               0.41               0.00
8     Seymour, I  P      2500000             253747      0.101              253747.0               9.73              -0.01
9  Mlodzinski, C  P       237000             247076      1.043              247076.0              -0.91              -0.03
10       Cruz, T SS       400000             231103      0.578              231103.0               0.69              -0.06
11     Warren, Z IF        30000             225755      7.525              225755.0              -0.92              -0.02
12     Gentry, T OF       460000             221233      0.481              221233.0              14.33              -0.02
13    Chavers, P CF      1500000             218867      0.146              218867.0               2.26              -0.01
14   Auerbach, B  C       125000             214380      1.715              214380.0              -0.92              -0.02
15     Bedell, I  P      2500000             202187      0.081              202187.0              19.00              -0.06
16    Workman, G SS       168500             200859      1.192              200859.0              -0.93              -0.01
17    Ramirez, B SS       343400             195418      0.569              195418.0               1.04              -0.03
18      Nahas, J  P        50000             191242      3.825              191242.0              -0.85              -0.02
19      Baker, D 2B       400000             230053      0.575              184042.4               7.00               0.20
20      Boyle, J  P      4000000             174688      0.044              174688.0               9.00              -0.24
21     Hauver, T OF        55000             164555      2.992              164555.0              -0.99              -0.06
22  Campbell, N 2B       600000             197337      0.329              157869.6               9.91               0.20
23    Dollard, T  P       160000             151411      0.946              151411.0              -0.73              -0.23
24    Hardman, T 3B       455600             183473      0.403              146778.4               1.85               0.21
25     Oviedo, A SS       336600             145856      0.433              145856.0              -0.26              -0.21
```

Above is the board I took with me into the draft. I had the 12th pick in the first round. I was skeptical Gonzales would fall, but confident Kavadas could fall. That is until the moderator of the draft, Wareham Gatemen GM Andy Lang, called each of us into his office individually before the draft. Andy said that there would be opportunities to agree to a premium value for a player before the draft, in exchange for the guarantee that you would get the player right there and then before the draft. One of the players was Nick Gonzales for $5 million in the 2nd round. My mind lit up like the Christmas tree, thought about it for 2 seconds, thought about the 30% premium, and said yes.

With Gonzales secured, I picked Kavadas in the first round. I had drafted the top two players on my board. I then turned my attention to pitching, given the steep decline in quality, and selected Mlodzinski and Nahas in the next two rounds. My final roster is below:

Round	Name	Position	School	Dollar Amount
1	Niko Kavadas	INF	Notre Dame	65,000
2	Nick Gonzales	2B	New Mexico St.	5,000,000
3	Carmen Mlodzinski	P	South Carolina	237,000
4	Joseph Nahas	P	Georgia Southern	50,000
5	Brett Auerbach	C	Alabama	125,000
6	Zavier Warren	INF	Central Michigan	30,000
7	Jacob Teter	1B	Florida Southern	500,000
8	Noah Campbell	2B	South Carolina	100,000
9	Noah Skirrow	P	Liberty	800,000
10	Jared DeSantelo	3B	Florida Atlantic	65,000
11	Austin Love	P	North Carolina	1,000,000
12	Reid Johnston	P	NC State	500,000
13	Andrew Abbott	P	Virginia	1,125,000
14	Lucas Dunn	IF/OF	Louisville	906,800
15	Bryce Osmond	P	Jenks High School	1,000,000

I know I stuck to my board when I ended up drafting 8 of my top 25 players on the board. I got the taste that the draft is a very human process. As much as I ponder a design of an algorithm to

encompass the whole selection process happening in our brains, there will always be humans to negotiate values, make the phone calls, and ultimately drive the operation. I found myself throughout the draft taking players further down my board for cost purposes. I used the model as a guide but did not stick to it strictly. In addition, I know my objective function was not all encompassing with respect to prediction, so I did not treat it as such. Lastly, things like character/makeup/and body type are things that must be taken into account and I wish I could have had the opportunity to build these things into a utility function. As we move towards a more complete and elaborate understanding of players, I look forward to building more complete models that are holistically encompassing and aid in the player evaluation and selection processes of the MLB Amateur Draft.