An Analysis of Rookie NFL Wide Receivers Using an NBD Model

Intro

If you are a general manager or sports bettor, rookie wide receivers in the National Football League (NFL) are some of the most volatile assets in the league. There is very little data about these players before the start of the regular season outside of their statistics at the collegiate level, but high expectations for what they can contribute to an offense. As such, a more robust model of wide receiver success at the rookie level can lead to a meaningful advantage for anyone invested in their success.

This report analyzes the performance of rookie wide receivers from a fantasy football perspective — an online game where players can simulate owning a football team (see “Background” below). In particular, we model the number of receptions (catches) rookie wide receivers get and consider the story behind three covariates to form a thesis about which rookie wide receivers have the best chance of standing out and hence, are currently undervalued.

We find that using more precise measurements of time *t* leads to better models and the number of pass attempts per team is the strongest covariate for receptions. The underlying analysis also highlights the variance and unpredictability associated with fantasy football, which passionate players ride on season after season.

Background

Fantasy football allows participants to choose (“draft”) NFL players for a team that they manage for an entire season. Typically, participants join a league with other participants and draft players in a round-robin format, where each player can only be on one team. NFL players are given points throughout the season based on their performance in actual NFL games, and the participant who wins the most games by consistently having the higher-scoring team wins the season.

This study uses data on drafted rookie wide receivers for the past ten seasons from 2014 to 2023, of which there were 253 rookies. The data was sourced from Pro Football Reference and the nflverse Github, and it was arranged using Python code. The processed dataset includes information about players, their receptions and touchdowns, as well as statistics about their team.

Player	Rnd	College/Univ	offense_snaps	Tm	G	Rec	TD	PF	Att	year
Sammy Watkins	1	Clemson	1026	BUF	16	65	6	343.0	579.0	2014
Mike Evans	1	Texas A&M	769	TAM	15	68	12	277.0	531.0	2014
Odell Beckham Jr.	1	LSU	771	NYG	12	91	12	380.0	607.0	2014
Brandin Cooks	1	Oregon St.	533	NOR	10	53	3	401.0	659.0	2014
Kelvin Benjamin	1	Florida St.	1049	CAR	16	73	9	339.0	545.0	2014
Marqise Lee	2	USC	492	JAX	13	37	1	249.0	557.0	2014
Jordan Matthews	2	Vanderbilt	765	PHI	16	67	8	474.0	621.0	2014
Paul Richardson	2	Colorado	511	SEA	15	29	1	394.0	454.0	2014
Davante Adams	2	Fresno St.	861	GNB	16	38	3	486.0	536.0	2014
Cody Latimer	2	Indiana	37	DEN	8	2	0	482.0	607.0	2014
Allen Robinson	2	Penn St.	516	JAX	10	48	2	249.0	557.0	2014
Jarvis Landry	2	LSU	683	MIA	16	84	5	388.0	595.0	2014

Figure 1: Sample of the full dataset used in the report.

Unless otherwise stated, statistics represent the aggregate statistics for the player across the entire season. As a quick overview, *Rnd* represents the round in which the player was drafted (one through seven). *Offense_snaps* are the number of offensive plays the player was on the field for. *Tm* contains three-letter abbreviations for the NFL team that the player was drafted to. *G* is the number of games in which the player was active, *Rec* is the number of receptions, and *TD* is the number of touchdowns. *PF* is the number of points scored by the player's team and *Att* is the number of pass attempts by the team. Finally, *year* represents when the player was drafted (2014 to 2023).

Only drafted wide receivers were factored into this study because, from a fantasy football perspective, a player should statistically never consider drafting a player who went undrafted in the actual NFL draft. Furthermore, undrafted NFL players will see a high variance in playing time with very few opportunities, which would make the data used in this study more noisy.

We choose to focus on receptions because the most common Fantasy Football scoring rule is Points-Per-Reception (PPR), which means wide receivers earn a point for every reception they have. Additionally, receptions are also a less variable metric: a player may go weeks without a touchdown, but they'll tend to have the same target share for receptions across different weeks.

Model Results

a. Generic Model

We first fit a generic NBD model with *G*, the games played, as the time unit. Although all teams play 16 regular season games, players will sometimes play less due to injuries. Therefore, we factor in the differences in *G* for each player by using the varying-*t* strategy.

The number of receptions by a rookie wide receiver over the past ten seasons ranges from zero to 105 (Puka Nacua in 2023). To organize the data, we chose to bin the receptions in groups of five (i.e. 0-4, 5-10, …, 70-74), with a right censor at 75+, similar to the task in HW2. Groups of five allow us to smooth over any natural fluctuations in data while preserving the underlying details; from a coaching perspective, having 30 versus 33 receptions over a season isn't a significant difference. There wasn't a scientific rationale for choosing five in particular besides the fact that it was a nice number and allowed most bins to have a count greater than 5.

Our generic model returned parameters $r = 1.34$ and $\alpha = 0.67$. Running the Chi-squared Goodness of Fit test returned a p-value of 0.026 — above the threshold of 0.001, but below what we would generally consider to be a good model.

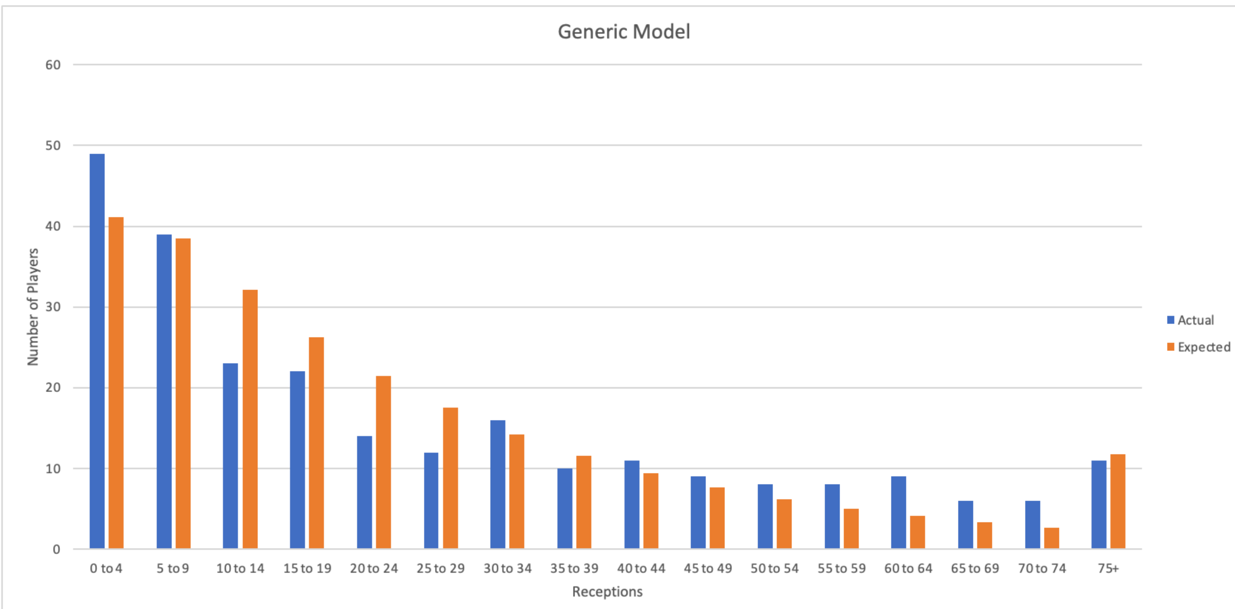

Figure 2: Generic model, histogram

Visually, we see that the generic model underestimated the number of players with 0 to 4 receptions and overestimated the number of players with 10 to 14, 15 to 19, and 20 to 24 receptions. More generally, the actual data rises and falls, while the NBD expected values tend to be more smooth.

The underestimate of the number of players with 0 to 4 receptions and subsequent overestimations would normally suggest a spike in the 0-4 receptions bin. Intuitively, this means that there is a proportion of players who are guaranteed to catch very few passes in the season — potentially due to injury, playing as a third-stringer, or being a poor wide receiver. However, running the solver on the model with a spike in the 0-4 bin led to $\pi = 0$ (i.e. the spike is not useful).

b. Model with Snaps as t

Using games as the time unit does not make much sense because different players will have different opportunities for receptions in each game. Under the generic model in a), a player who got injured in the first quarter versus a player who played all four quarters is treated the same on our time scale. For a more precise measurement, using the number of offensive snaps that each player plays would work better.

To do this, we refit the model using the *offense_snaps* column of the dataset, which returned parameters $r = 8.28$ and $\alpha = 133.6$. This is a large difference from the parameters in the generic model. In particular, we see that the players are more homogenous than initially presumed in our model with more precise time. The larger value for α also makes sense since we have modeled the data on a smaller time scale (number of snaps versus number of games).

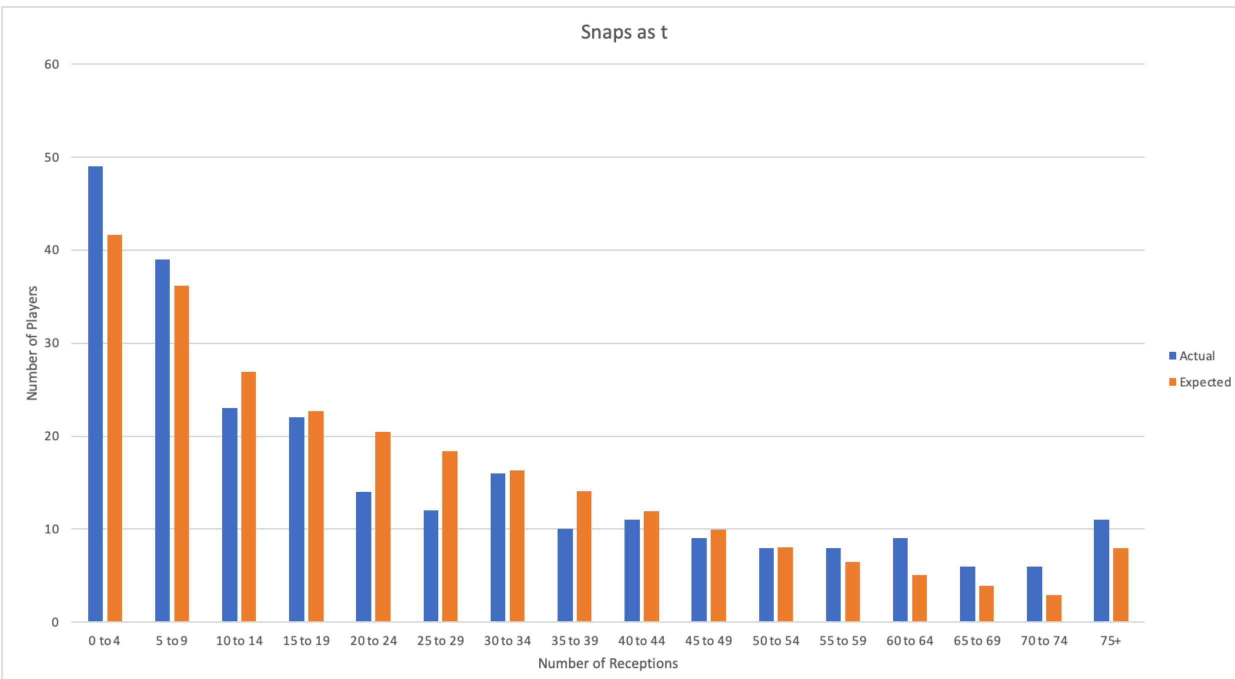

Figure 3: Model using snaps with t, histogram

Like the generic model, this model also underestimates the 0 to 4 bin and overestimates other bins from 10 receptions to 29 receptions. While the same intuition behind fitting a spike at the 0-4 bin holds as in the generic model, we again find that the spike is not useful with $\pi = 0$.

Note, however, that despite a better fit, this data is less useful to a coach than the generic model. This is because the number of snaps that a wide receiver will play is unknown before the start of a game, while whether or not they are going to play is public information. Unless we have a way to effectively model the number of snaps a player will play, this model does not provide us with much foresight to make a good coaching decision. Thus, we only use this model to gain additional insight into the information that we don't have available, and we stick to using games as our time scale for the rest of the models.

c. Covariates

We now consider the effects of three covariates included in the dataset: draft round, points scored by team, and passing attempts by team.

We consider draft rounds because players drafted in earlier rounds are considered to be better NFL prospects than those drafted later on. Hence, they are expected to be better wide receivers too, including in their rookie year. Furthermore, a team using an early-round draft pick on a wide receiver suggests that they're looking to focus on that department of their offense, so a receiver picked in the earlier round might have more opportunities than those picked later.

We partition the players into early-round draft picks (rounds 1 to 3) and later-round draft picks (rounds 4 to 7). Fitting a model to the early-round draft picks yields $r = 2.298$ and $\alpha = 0.863$. The p-value of the model after running the Chi-squared Goodness of Fit test is 0.03, which should be taken with a grain of salt since the partitioned dataset is much smaller.

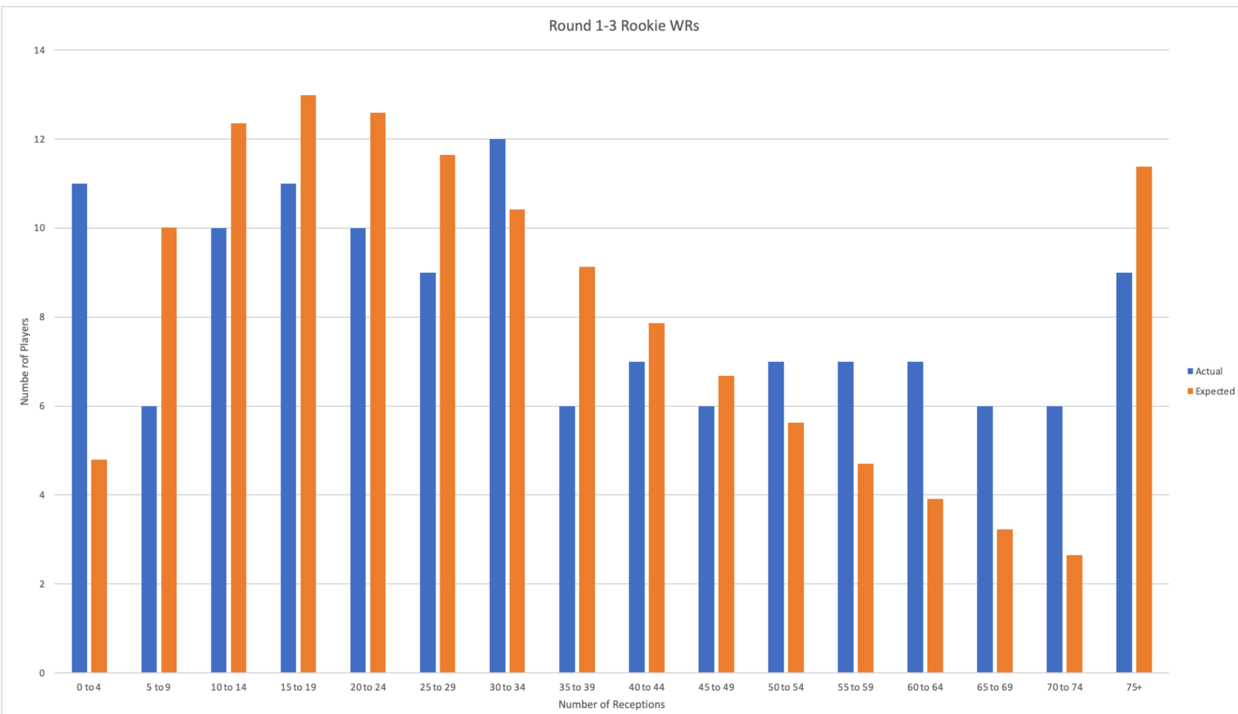

Figure 4: Early-Round wide receivers, histogram

Bin (Receptions)	Number of Players
0 to 4	11
5 to 9	6
10 to 14	10
15 to 19	11

20 to 24	10
25 to 29	9
30 to 34	12
35 to 39	6
40 to 44	7
45 to 49	6
50 to 54	7
55 to 59	7
60 to 64	7
65 to 69	6
70 to 74	6
75+	9

Table 1: Bin counts for early-round wide receivers

The visual representation of the model shows that it performs quite poorly, but also that the actual data seems to have an interesting pattern. This is highlighted by Table 1. While the bin of 30 to 34 receptions has the most players, it appears that otherwise, the number of receptions is more evenly distributed across the remaining bins and doesn't right-tail towards zero as we see in other count datasets.

For later-round wide receivers, we have $r = 1.17$ and $\alpha = 0.913$, with a p-value of 0.0014. This suggests that these wide receivers are more heterogenous, which is also reflected visually.

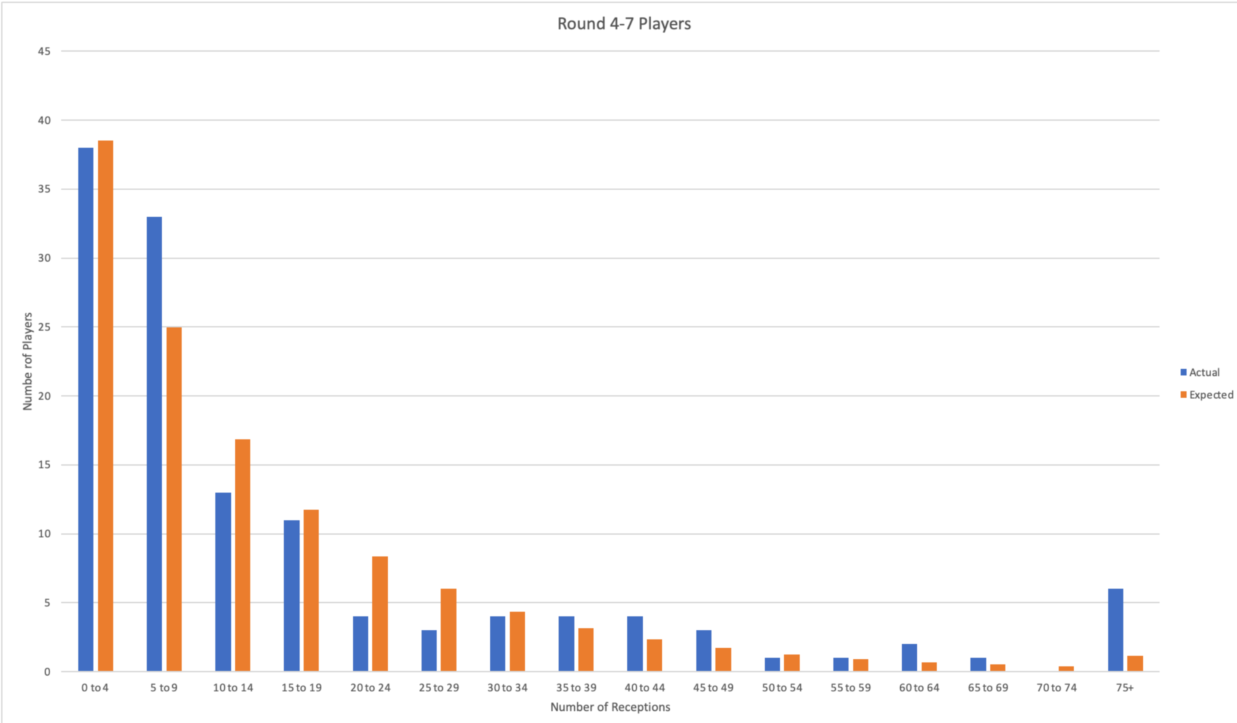

Figure 5: Later-Round wide receivers, histogram

Next, we consider the impact of the points scored by a player's team, or *Pf*. A higher number of points scored suggests that the team has a stronger offense, which could correlate to better play or more opportunities for a wide receiver. We base the model on the points scored on average per game since NFL teams only began playing 16 regular season games in 2021. As with the first covariate, we partition based on the top and bottom 50% of points scored per game.

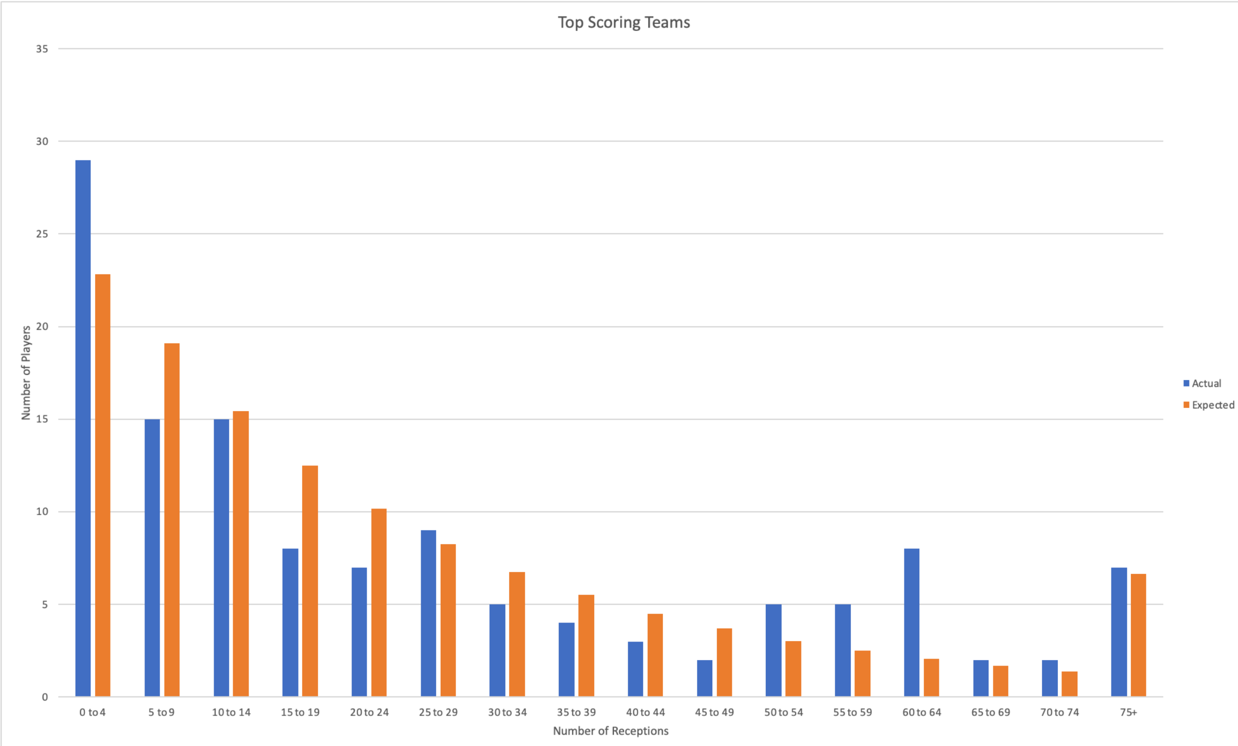

Figure 6: Top scorers, histogram

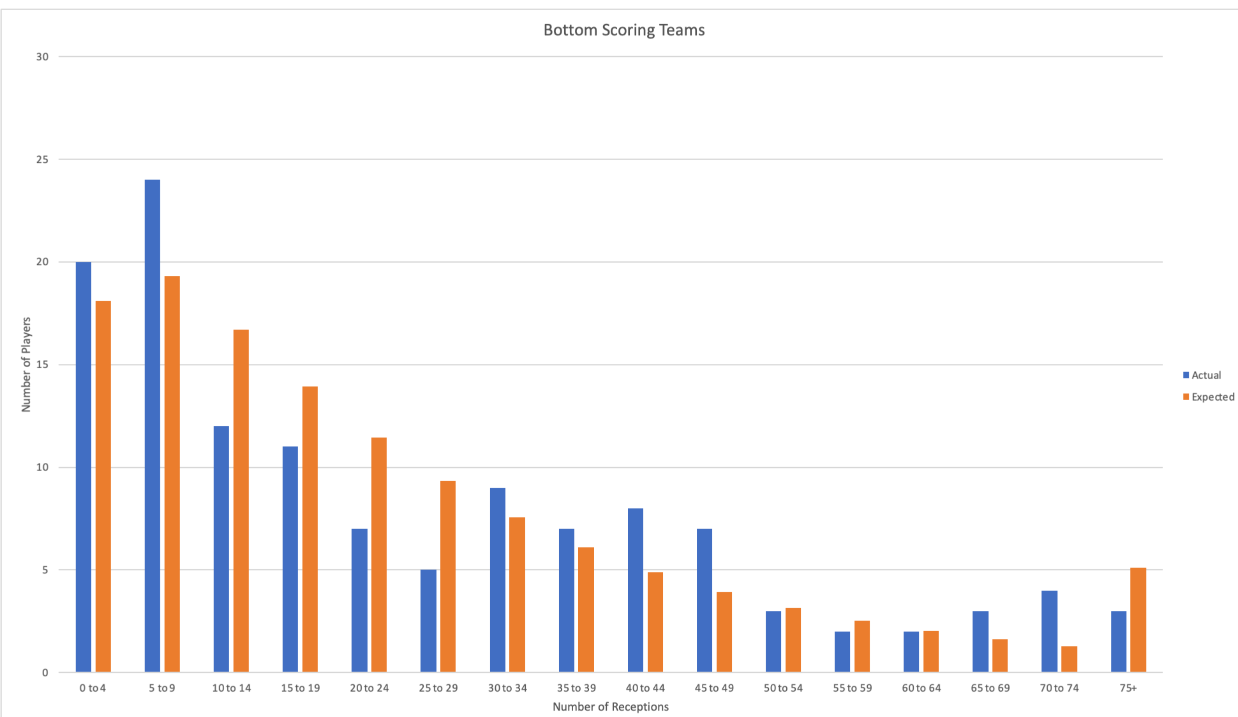

Figure 7: Bottom scorers, histogram

The parameters for the top scorers and bottom scorers are $r = 1.19$, $\alpha = 0.6$ and $r = 1.52$, $\alpha = 0.77$ respectively. We see visually that Figure 6, the top scorers, fits similarly to our generic model and has similar r, α values as well. However, both models have large overestimations and underestimations at certain points and don't meet our threshold p-value of 0.2.

Finally, we consider the number of passing attempts as a covariate. Like the points scored, we break this into the average passing attempts per game and partition it into the top 50% and the bottom 50%.

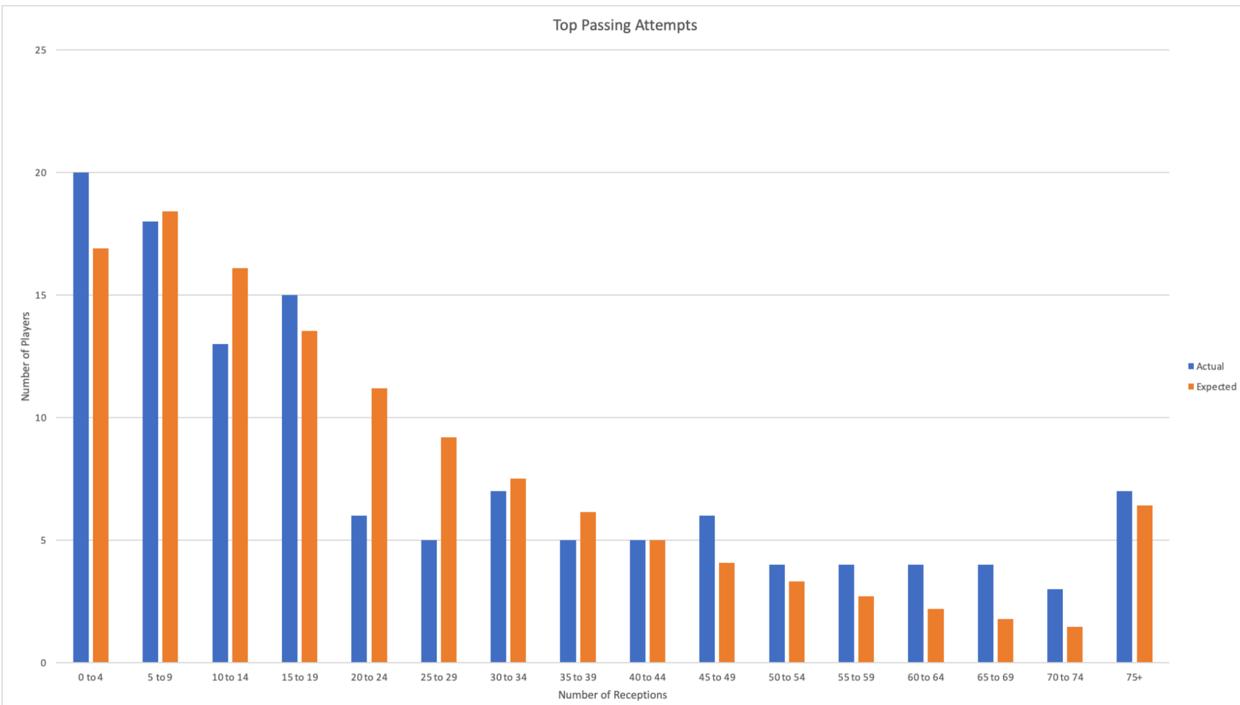

Figure 8: Top passing attempts, histogram

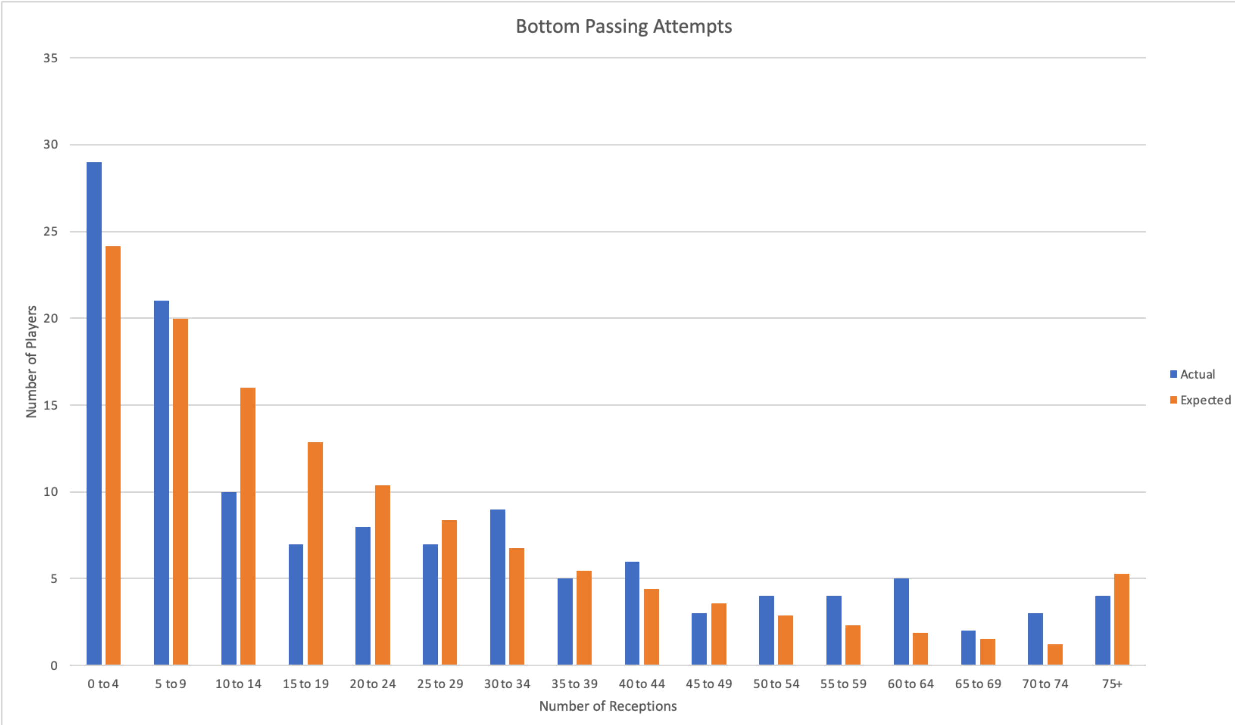

Figure 9: Bottom Passing Attempts, histogram

Intuitively, passing attempts should be a strong covariate because this directly leads to more opportunities for a wide receiver to catch passes. Indeed, we find that this covariate yields the highest p-values out of the three covariates considered, with parameters for the top passers and bottom passers being $r = 1.43$, $\alpha = 0.68$, p-value $= 0.412$ and $r = 1.26$, $\alpha = 0.68$, p-value $= 0.16$ respectively.

Takeaways

There was a significant difference between the models using games (generic model) and snaps (snaps model) as a unit of time, not only in how well the model fit, but also in how heterogenous the model was. Our generic model suggested that the wide receivers were almost heterogeneous, while the snaps model showed much higher homogeneity, which we should trust more because of its increased precision. This highlights two things:

1. Rookies who have a high number of receptions tend to be more lucky than skillful.
2. Using information that can only be learned after a game gives us much more power than before.

Practically, point one makes sense. A wide receiver's first year in the NFL depends on many more things than his skill, such as his team's offensive scheme or quarterback. Point two highlights the unpredictability of fantasy football: no matter how hard you try, you will most likely be shooting an arrow into the dark because there is important data that you simply cannot access in the present. Unfortunately, there's only so much one can do to get better at shooting an arrow in the dark compared to their opponents, and so much of the game still lies on variance.

Studying the covariate data also revealed different tendencies among the partitions we created. Later-round wide receivers were less homogenous than early-round wide receivers. Therefore, a knowledgeable fantasy football player can obtain a larger advantage by focusing on later-round wide receivers than earlier-round ones, such as by identifying the factors that could set such a player apart.

Surprisingly, the partitions for team scoring were only slightly homogenous or in other words, wide receivers on good offenses are not created equal. Hence, fantasy football players should not blindly draft a receiver on the Kansas City Chiefs, as there exists a deeper nuance that needs to be understood beyond being defending Super Bowl champions. This makes sense, as we see below that team pass attempts were the best covariate.

The partitions for team pass attempts were also only slightly homogenous, but they presented the best fit for the actual data in terms of p-values. These would be the best covariate option to use if someone wanted to project the probabilities of a player having a number of receptions. As above, when drafting, the lack of strong homogeneity suggests that wide receivers on high-passing offenses are also not created equal.

One more interesting note from the models was that the spikes placed at the 0-4 bin were not effective. This was unexpected since looking visually at the model, we see that the number of players with 0 to 4 receptions are underestimated and bins after are overestimated. However, from a story level, this makes sense. We chose to disregard wide receivers that were not drafted from our data, and these players would be the spiked population since they would only be activated to serve as backups or they would play poorly.

In general, the data suggests that projecting wide receiver receptions is more luck-based than not, but not homogenous enough that there doesn't exist a method to the madness. Fantasy football players, general managers, and sports bettors can leverage the differences in heterogeneity to make decisions that most leverage the perceived edge they have.

Areas of Improvement

A limitation in the robustness of the study was the amount of data used. While 253 players were enough for the generic model, they bottlenecked the number of segments that could be formed off of covariate data. In particular, a better segmentation for the round drafted covariate would consist of rounds 1-2, 3-5, and 6-7, since there's a relatively heavy drop-off in a player's quality from round 1 to round 3 and 4 to 7. Using more, and better, segments would unlock even more insight into how heterogeneity differs across the different partitions.

The model also assumes that a player remains with their drafted team for the entire duration of their season. Generally, this is not an unrealistic assumption, since rookie players are given time to develop and are less likely to be traded or cut. However, it does diminish precision in the data used in covariates two and three. The parameters would likely be very slightly different if we took this into account.