Country Over Club?: An Analysis of USWNT Players and NWSL Club Performance

GR5015 Final Project

Christa Chiao (cjc2297)

December 14, 2021

Introduction

Overview

This study looks to investigate the impact of United States Women's National Team (USWNT) players on their respective club teams' performance in their domestic league, the National Women's Soccer League (NWSL). USWNT players play for both their club team and their national team, often missing significant swaths of regular club seasons to play in international tournaments, friendlies, or to train in national team camp. The NWSL has historically not stopped their play during these major tournaments with teams forced to play their younger or less experienced players in starting slots. This can impact the overall performance of these club teams in numerous ways: younger or less experienced players can make significant mistakes, causes teams to lose games they would have likely won or these players can have a breakthrough moment: a twenty-one year old rookie has an absolutely beautiful goal, or a goalkeeper has the shutout of their life. The number of national team players that a club keeps on their roster is a matter of debate: is it more advantageous (purely from a team performance perspective) to have a roster of players who are able to have full availability, or to have one or two star players? This study endeavors, in part, to answer this question.

Research Question and Hypothesis

This study poses the question:

Do USWNT players impact their club team performance during the 2016 through 2019 regular seasons? What is the impact of US-allocated players on club performance, specifically, games won, games drawn, games lost, goals for, playoff qualification, and league standing?

My hypothesis is that USWNT players starting or coming off the bench during regular season NWSL games has a meaningful relationship to the performance of club teams. I believe that the more often USWNT players are involved in regular season games, the better their club team's performance, which will be defined in the Variables section and the Data Description.

Variables

The independent variable for this study is the frequency that USWNT players are either starting or coming off the bench during NWSL regular season games. This includes both field players and goalkeepers. For the purposes of this study, we will define "USWNT players" as professional athletes who are federation players, as indicated by the United States Soccer Federation.[1]

[1] Federation players, or "allocated players", are professional soccer players whose salaries for playing in the NWSL are paid for by their national federations. In the case of this study, that is the United States Soccer Federation. Source is Jeff Kassouf at Equalizer Soccer

season	team_name	wins	draws	losses	goals_for	league_ranking	playoffs
2016	Boston Breakers	3	2	15	14	10	0
2016	Chicago Red Stars	9	6	5	24	3	1
2016	Houston Dash	6	4	10	29	8	0
2016	FC Kansas City	7	5	8	18	6	0
2016	Orlando Pride	6	1	13	20	9	0
2016	Portland Thorns FC	12	5	3	35	1	1

The dependent variable is the performance of club teams during the regular season. This project will define "club team performance" with the following:

- **Games Won**: The number of games won during the regular season
- **Games Drawn**: The number of games that ended in a draw during the regular season
- **Games Lost**: The number of games that ended in a loss during the regular season
- **Goals For**: The number of goals scored by the club during the regular season
- **Playoff Qualification**: This is a binary 1 or 0 variable where 1 is that the club qualified for the playoffs and 0 is that the club did not qualify for the playoffs
- **League Standing**: The standing of the clubs where 1 indicates that the club was first in the league standings at the end of the regular season

Scope of Study

This study will limit its scope to the 2016, 2017, 2018, and 2019 regular seasons as to include a broad enough swath of games and types of years, but will limit to games and seasons prior to the COVID-19 pandemic, which changed the structure of the season for the NWSL, culminating in the Challenge Cup. This study will also limit its scope to the players allocated by the United States Soccer Federation. The NWSL contained allocated players from three different countries in 2016 - 2019: Canada, Mexico, and the United States.

This project does not happen without the creation of the `nwslR` package by Arielle Dror and Sophia Tannir. Thank you to these pioneers for creating and investing in women's soccer analytics.

Data Description

Data Set

This study will be working with two data sets primarily. The first and primary data set was scraped from the NWSL website by the creators of the `nwslR` package. The data housed in the `nwslR` package include ID tables and statistics. We will be using the `nwslR::team_stats_season` table. We have abbreviated this table to include data that is within the scope of the study, from 2016 to 2019, and narrowed the data to include just the metrics that this study is interested in working with, as well as identifying data points, including the season the team played, the team_id, and performance metrics. League ranking per team per season was manually combined and added as a csv.

The second data set that this study will use is a player-level data set. This data set is limited to field-players, or players who do not play in goal. This study will exclude goalkeepers. In order to identify individual players, we have joined in the `player` data set, which includes player name, the position on the field the player plays,

player_name	season	pos	team_name	mp	starts
Alex Arlitt	2016	DF	FC Kansas City	11	8
Yael Averbuch	2016	MF	FC Kansas City	20	20
Lauren Barnes	2016	DF	Seattle Reign FC	20	20
Lauren Barnes	2016	DF	Reign FC	20	20
McKenzie Berryhill	2016	DF	Portland Thorns FC	5	2
Celeste Boureille	2016	MF	Portland Thorns FC	6	5

player_name	season	pos	team_name	allocated
Alex Arlitt	2016	DF	FC Kansas City	0
Yael Averbuch	2016	MF	FC Kansas City	0
Lauren Barnes	2016	DF	Seattle Reign FC	0
Lauren Barnes	2016	DF	Reign FC	0
McKenzie Berryhill	2016	DF	Portland Thorns FC	0
Celeste Boureille	2016	MF	Portland Thorns FC	0

and any aliases the player might go by. Additionally, we've limited the data set to what is in the scope of this study: filtering to players from the United States and to players within the 2016 - 2019 seasons.

We also need to figure out which players are allocated from the years 2016 - 2019. This is not a publicly available pre-compiled data set (that I could find!) so I manually searched and compiled this data set.[2] We are joining this data set to the `player_stats` data set from above with a binary flag for `allocated == 1` where the player was a United States Federation allocated player that year and `allocated == 0` if the player was *not* a USF allocated player that year. With that final addition, we have our final, person/player-level data set.

One final step is to join these two data sets together to have our final data set with

Variables

The independent variable of the study is if an NWSL player is allocated by the US Soccer Federation (USSF), which is a binary value in the `player_stats` data set. These data were collected from a number of sources online. Which players USSF chooses to allocate and how many players USSF chooses to allocate varies from year to year, although there is some consistency. The USSF tends to semi-consistently allocate players, although the club teams these players play for are not always consistent, and the number of allocated players on each team varies over the seasons.

The dependent variables are as follows:

- **Games Won**: The "Games Won" variable is the number of games won during the regular season.

[2] Sources for the allocation draft are hyperlinked by year: 2016, 2017, 2018, 2019

season	team_name	goals_for	wins	draws	losses	league_ranking	playoffs	total_alloc	mean_min
2016	Boston Breakers	14	3	2	15	10	0	1	1260
2016	Chicago Red Stars	24	9	6	5	3	1	2	1170
2016	Houston Dash	29	6	4	10	8	0	2	779
2016	FC Kansas City	18	7	5	8	6	0	2	1233
2016	Orlando Pride	20	6	1	13	9	0	1	1350
2016	Portland Thorns FC	35	12	5	3	1	1	4	1268

season	0	1
2016	133	20
2017	143	25
2018	130	22
2019	157	23

season	wins_mean	draws_mean	losses_mean	goals_for_mean
2016	7.91	4.36	7.73	26.55
2017	9.45	5.18	9.36	35.18
2018	8.90	6.80	8.30	30.50
2019	9.50	5.40	9.10	30.90

This number is represented as `wins` in the `team_stats` data set. This number is unchanged from the raw data and ranges from 1 to 17, with a median of 7 wins and has a standard deviation of 3.721.

- **Games Drawn**: The "Games Drawn" variable is the number of games that ended in a draw during the regular season. This number is represented as `draws` in the `team_stats` data set. This number is unchanged from the raw data and ranges from 1 to 10 with a median of 5 draws and has a standard deviation of 2.002.
- **Games Lost**: The "Games Lost" variable is the number of games that ended in a loss during the regular season. This number is represented as `losses` in the `team_stats` data set. This number is unchanged from the raw data and ranges from 1 to 17 with a median of 8 losses and has a standard deviation of 3.962.

- **Goals For**: The "Goals For" variable is the number of goals scored by the club during the regular season. This number is represented by the `goals_for` variable in the `team_stats` data set. This number is unchanged and unprocessed from the raw data and ranges from 12 to 54 with a median of 29.5 goals scored for a team. The standard deviation of `goals for` is 9.959.

- **League Standing**: The "League Standing" variable is the place in which clubs were on the final day of the regular season. This variable was gathered separately from the `nwslR` package and joined to the data. The data is organized where if a team has a `league_ranking == 1`, they were first in the league on points earned during the regular season.[3]

- **Playoff Qualification**: The "Playoff Qualification" variable is a simplification of the "League Standing" variable, and is a binary variable where 1 is that the club qualified for the playoffs and 0 is that the club did not qualify for the playoffs during the regular season. This variable is imputed from "League Standing" as teams automatically qualify for playoffs if they have a league standing of 4 or lower (which is higher) at the end fo the regular season.

Descriptive Statistics

I broke out our variables of interest into a number of tables on two different variables: by **season** and by **team**. I further divided the descriptive statistics into game-specific statistics (Games Won, Games Drawn, Games Lost, and Goals For) and season-round-up variables (League Ranking and PlayofF Qualification).

As you can see in the table above, the mean win number across the seasons is fairly consistent and interacts

[3]In the NWSL (and broadly across all soccer leagues), teams will earn 3 points for every game they win, 1 point for every draw, and 0 points for every game they lose. If teams are tied on points, the places are decided on goal difference, which is the number of goals that the team has scored over the course of the regular season subtracting the number of goals that the team has conceded over the course of the regular season. If teams are tied on points after this, goals for is the deciding variable.

season	wins_sd	draws_sd	losses_sd	goals_for_sd
2016	2.63	1.69	3.64	7.54
2017	3.56	2.14	3.38	7.87
2018	4.65	1.62	5.27	11.56
2019	3.50	2.01	3.78	10.80

team_name	wins_mean	draws_mean	losses_mean	goals_for_mean
Boston Breakers	3.50	4.500	14.000	19.000
Chicago Red Stars	10.75	6.000	6.250	34.000
FC Kansas City	7.50	6.000	8.500	23.500
Houston Dash	7.25	4.250	11.500	27.000
North Carolina Courage	16.00	3.667	4.333	48.333
Orlando Pride	7.25	4.500	11.250	29.750
Portland Thorns FC	12.25	5.750	5.000	38.000
Reign FC	9.50	7.250	6.250	31.500
Seattle Reign FC	9.50	7.250	6.250	31.500
Sky Blue FC	5.75	4.750	12.500	26.750
Utah Royals FC	9.50	6.000	8.500	23.500
Washington Spirit	7.00	4.750	11.250	25.500
Western New York Flash	9.00	5.000	6.000	40.000

with the mean draw number and the mean loss number as the mean wins and mean losses decreases, the mean draws increase. The mean number of goals peaked in 2017, with 35.182 goals for each team.

The following table breaks out game statistics by team. The Boston Breakers were dissolved in 2017 and the Western New York Flash was dissolved in 2018.

As can be seen in the table above, the North Carolina Courage leads the teams in the mean number of wins with 16 wins, with Portland Thorns FC and Chicago Red Stars in second and third with 12.25 wins and 10.75 wins, on average, respectively. The Boston Breakers lead the average losses with 14.000 losses, followed by Sky Blue FC with 12.500 losses. The North Carolina Courage leads the league in average number of goals with 48.333, followed by the Western New York Flash (40.000) and Portland Thorns FC (38.000).

This table shows the average league position in `league_ranking_mean` as well as the percentage success that a club has had reaching the playoffs (in the number of attempts).

team_name	wins_sd	draws_sd	losses_sd	goals_for_sd
Boston Breakers	0.707	3.536	1.414	7.071
Chicago Red Stars	2.363	3.266	1.500	7.439
FC Kansas City	0.707	1.414	0.707	7.778
Houston Dash	1.258	0.957	1.915	6.325
North Carolina Courage	1.000	2.517	3.055	8.963
Orlando Pride	2.986	2.646	4.272	10.966
Portland Thorns FC	1.258	0.957	1.414	2.449
Reign FC	1.291	0.957	1.258	7.724
Seattle Reign FC	1.291	0.957	1.258	7.724
Sky Blue FC	3.775	1.258	3.873	10.308
Utah Royals FC	0.707	2.828	2.121	2.121
Washington Spirit	4.397	1.708	5.679	9.000
Western New York Flash	NA	NA	NA	NA

team_name	league_ranking_mean	league_ranking_sd	playoff_mean	playoff_sd
Boston Breakers	9.50	0.707	0.00	0.000
Chicago Red Stars	3.25	0.957	1.00	0.000
FC Kansas City	6.50	0.707	0.00	0.000
Houston Dash	7.25	0.957	0.00	0.000
North Carolina Courage	1.00	0.000	1.00	0.000
Orlando Pride	7.00	2.828	0.25	0.500
Portland Thorns FC	2.00	0.816	1.00	0.000
Reign FC	4.25	0.957	0.50	0.577
Seattle Reign FC	4.25	0.957	0.50	0.577
Sky Blue FC	7.50	1.291	0.00	0.000
Utah Royals FC	5.50	0.707	0.00	0.000
Washington Spirit	6.25	3.500	0.25	0.500
Western New York Flash	4.00	NA	1.00	NA

team_name	mean	sd
Boston Breakers	0.069	0.258
Chicago Red Stars	0.154	0.363
FC Kansas City	0.152	0.364
Houston Dash	0.098	0.300
North Carolina Courage	0.216	0.415
Orlando Pride	0.089	0.288
Portland Thorns FC	0.224	0.420
Reign FC	0.109	0.315
Seattle Reign FC	0.109	0.315
Sky Blue FC	0.078	0.270
Utah Royals FC	0.229	0.426
Washington Spirit	0.121	0.329
Western New York Flash	0.154	0.376

The North Carolina Courage, Portland Thorns, and the Chicago Red Stars lead the mean League Ranking variable (by having the lowest values), while the Boston Breakers, Sky Blue FC, and the Houston Dash are at the bottom of the League Ranking variable (by having the highest values). In the four regular seasons that make up the 2016-2019 years, the Chicago Red Stars, North Carolina Courage, and the Portland Thorns have qualified for all four playoff tournaments[4], Seattle Reign FC qualified for 50%, or two of four playoff tournaments, the Washington Spirit and the Orlando Pride qualified for 25%, or one of four playoff tournaments, and the Boston Breakers, FC Kansas City, Houston Dash, Sky Blue FC, and Utah Royals have qualified for zero playoff tournaments between 2016 and 2019.

This allocation table indicate the mean and standard deviation of the USSF-allocated number of players.

Initial Models

Tell me what model you are using and why (logit, probit, LPM, fixed effects, etc.). Start off with a simple model relating you main IV to your main DV. Explain the relationship and why this initial model is insufficient. Maybe you need to make a scale/index of variables. Maybe you need to control for additional factors. Maybe you want to include interaction terms. Maybe you need to check for serial correlation. Etc. Interpret everything correctly (ceteris paribus, on the right scale, etc.)

[4]The Western New York Flash qualified for playoffs in the 2 seasons they existed in this study: 2017 and 2018.

I'm initially starting with a multiple linear regression models to predict league ranking and the number of wins that a team has. The first multiple linear model is moderately successful in understanding the relationship between the three allocated player variables (`total_alloc`, `mean_min`, and `mean_starts`). This prediction model looks like:

$$\text{LM1} = a + b_1\textbf{total\_alloc} + c_1\textbf{mean\_min} + d_1\textbf{mean\_starts}$$

The first multiple linear regression model, `lm1` shows that a unit increase of the `total_alloc` variable would, on average, increase the number of wins by 0.712 games, net of other variables. The `total_alloc` variable has a p-value well below .05 of 0.00703, and thus, at a 95% confidence interval is a statistically significant predictor of the number of wins that a club will have in a given season. The mean number of minutes, indicated by variable `mean_min` shows that with a unit increase in the number of minutes a USSF-allocated player has on the field, the team will have an increase of 0.021 wins, with a p-value of .125, on average and net of other variables. This indicates that the mean number of minutes a USSF-allocated player has on the field does not have strong correlation with the number of games won. The third variable, `mean_starts` shows an interesting turn: on average, with a unit increase in the number of starts that an allocated player has during the NWSL regular season, the team will win -1.740 games fewer, with a p-value = .163, net of other variables. This indicates that the mean number of starts by a US allocated player is not correlated with the number of games won and is a weak predictor of this metric.

I then tried to add the interaction term of `season` that I hypothesize will mediate the relationship between the number of players allocated to a team and the number of wins.

The second multiple linear regression model, `lm3`is not very effective. It shows that a unit increase of the `total_alloc` variable would, on average, increase the number of wins by 0.0028 games, net of other variables. The `total_alloc` variable has a p-value of 0.584 and thus, at a 95% confidence interval, is not a statistically significant predictor of the number of wins that a club will have. The mean number of minutes shows that with a unit increase in the number of minutes a USSF-allocated player has on the field, the team will have an increase of 0.0012 wins, with a p-value of .389, on average and net of other variables. This indicates that the mean number of minutes a USSF-allocated player has on the field does not have strong correlation with the number of games won. The third variable, `mean_starts` reports a relationship that is consistent with the first linear model: on average, with a unit increase in the number of starts that an allocated player has during the NWSL regular season, the team will win -0.0940 games fewer, with a p-value = .448, net of other variables. This indicates that the mean number of starts by a US allocated player is not correlated with the number of games won and is a weak predictor of this metric.

A table comparing the these multiple linear models can be found here (See Table 1: Model Comparison):

```r
stargazer::stargazer(lm1, lm2, type = "latex", title = "Model Comparison")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Dec 14, 2021 - 21:48:42

The second multiple linear regression model I'm running looks to predict the relationship between the overall league ranking and the three allocated player variables. This prediction model looks like:

$$\text{LM2} = a + b_1\textbf{total\_alloc} + c_1\textbf{mean\_min} + d_1\textbf{mean\_starts}$$

The third linear regression looks to investigate the relationship between a team's league ranking on the final day of the regular season and any allocation variables. This model shows that with a unit increase in the `total_alloc` variable, the team would, on average, place -1.194 places lower, net of other variables. This means that with an increase of one US allocated player, the league ranking would decrease by over a place. This p-value is well below .05, and the relationship is statistically significant at a 95% confidence interval. On average, a unit increase in the mean number of minutes that a US-allocated player plays during the regular season, net of other variables, causes a 0.01 decrease in league placement at the end of the regular season, although this is not a statistically significant variable. A unit increase in the number of starts by

Table 1: Model Comparison

	Dependent variable:	
	wins	
	(1)	(2)
total_alloc	1.456^{***}	283.467
	(0.360)	(724.635)
season		0.518
		(0.935)
mean_min	0.011	0.012
	(0.013)	(0.014)
mean_starts	−0.860	−0.940
	(1.187)	(1.226)
total_alloc:season		−0.140
		(0.359)
Constant	4.097^{***}	−1,041.618
	(1.298)	(1,886.770)
Observations	42	42
R^2	0.381	0.387
Adjusted R^2	0.332	0.302
Residual Std. Error	2.907 (df = 38)	2.972 (df = 36)
F Statistic	7.797^{***} (df = 3; 38)	4.548^{***} (df = 5; 36)
Note:		*p<0.1; **p<0.05; ***p<0.01

an allocated player results in a 0.965 increase in league placement, although again, this is not a statistically significant variable.

I then tried adding in the season as a variable because the games take place is an important factor in determining league placement, and should be kept in suspension. The results were similar to my other regressions, with the number of allocated players as significant in relation to a team's league ranking. On average, a unit increase in the total number of US allocated players would cause a 1.190 decrease in league placement, net of other variables. The p-value is .0001.

A table comparing the multiple linear models that attempted to predict the league ranking can be found here (see Table 2: Model Comparison):

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Dec 14, 2021 - 21:48:42

Table 2: Model Comparison

	Dependent variable:	
	league_ranking	
	(1)	(2)
total_alloc	-1.194^{***}	-1.189^{***}
	(0.264)	(0.275)
season		-0.022
		(0.309)
mean_min	-0.011	-0.011
	(0.010)	(0.010)
mean_starts	0.965	0.967
	(0.868)	(0.880)
Constant	8.716^{***}	52.821
	(0.950)	(623.319)
Observations	42	42
R^2	0.415	0.415
Adjusted R^2	0.369	0.352
Residual Std. Error	2.126 (df = 38)	2.155 (df = 37)
F Statistic	8.999^{***} (df = 3; 38)	6.574^{***} (df = 4; 37)
Note:		*p<0.1; **p<0.05; ***p<0.01

Adding in the `season` variable to the second multiple linear model in the table did not hugely change coefficients or the result of the model. This is likely because while the season is an important indicator of time as well as shows important inflection points for how many or which allocated players play on which team, they are largely more consistent than I previously thought.

I decided my final move would be to add team name to the model. This model resulted in interesting, but ultimately unhelpful results. The model pulled out teams that have consistently shown good results and who have a sizable number of allocated players. I'll summarize only a number of these results. When adding in `team_name` into the model, the number of allocated players does not stand out as a statistically significant predictor of the number of wins that a team will have at the end of the regular season. The model indicates that with a unit increase in the number of allocated players, there will be, on average, a 0.420 increase in wins at the end of the regular season, net of other factors (p = .330). Being the Chicago Red Stars causes, on average, a 6.411 increase in the number of wins at the end of the regular season, net of other factors, and being the North Carolina Courage causes, on average, an 11.381 increase in the number of wins at the end of the regular season, net of other factors. These relationships are statistically significant at a 95% confidence interval, if not a helpful explanatory model. (See Table 3:Model 5)

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Dec 14, 2021 - 21:48:43

Final Models

This table contains the best models from my previous iterating:

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Dec 14, 2021 - 21:48:43

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Dec 14, 2021 - 21:48:43

My fourth and sixth models were my best models because of their relatively low RSE and their significant p-values at a 95% confidence interval. The fourth multiple regression model I ran was the best model for predicting the relationship between the number of USSF-allocated players and the league ranking a team would be at the end of the regular season, and the sixth model I ran was the best model for predicting the relationship between the number of USSF-allocated players and the number of games a team would win at the end of the regular season. (See Tables 4 and 5)

Conclusions

According to the statistical findings in this study, the increased number of USSF-allocated players will increase the number of regular season games that a team will win. Th This study showed a significant relationship between wins and the the number of USSF-allocated players on make up a roster. The number of minutes played by an allocated player, or the number of games where an allocated player is in the starting lineup does not have any bearing on the number of games a team wins, and the ultimate league ranking at the end of the regular season.

This analysis is quite limited to the data available and the time available to conduct this analysis. One limitation is noted at the top of this analysis: this study is limited to field players only, and the two to three USSF-allocated goal keepers are not included in the counts or various metrics in this study. This study also only limits to US players, while players in the NWSL include Canadian players and for a short while, Mexican players. Expanding the scope to include all of these groups would increase the strength of this study.

If I were to revisit this project, I would seek out more data than exists in the `nwslR` package. The NWSL was founded in 2012 and has run consistently, with the exception of 2020[5], through the 2021 regular season. I would gather as much data from the 9 years of the NWSL, as well as previous iterations of the NWSL, Women's Professional Soccer (WPS), which ran from 2007 until 2012, and the Women's United Soccer Association (WUSA), which ran from 2001 until 2003. This additional data would reduce error and allow further insights into the impact of USSF allocated players on women's club soccer performance.

With more time (and knowledge!), I would do a time-series analysis in order to see the impact of US-allocated players over the course of a season. I believe there would be really interesting insights to gather from these analyses.

[5] In 2020, the NWSL hosted the first inaugural Challenge Cup, which brought teams together into a tournament-style series of games that eventually crowned the Houston Dash champions. The NWSL was the first professional league in the world to implement a "bubble" system in order to continue play.

Table 3: Model 5

	Dependent variable:
	wins
total_alloc	0.420
	(0.424)
as.factor(team_name)Chicago Red Stars	6.411***
	(2.246)
as.factor(team_name)FC Kansas City	3.370
	(2.485)
as.factor(team_name)Houston Dash	3.645*
	(2.083)
as.factor(team_name)North Carolina Courage	11.381***
	(2.467)
as.factor(team_name)Orlando Pride	3.645*
	(2.083)
as.factor(team_name)Portland Thorns FC	7.596***
	(2.384)
as.factor(team_name)Reign FC	5.161**
	(2.246)
as.factor(team_name)Seattle Reign FC	5.161**
	(2.246)
as.factor(team_name)Sky Blue FC	2.145
	(2.083)
as.factor(team_name)Utah Royals FC	4.741*
	(2.718)
as.factor(team_name)Washington Spirit	3.080
	(2.123)
as.factor(team_name)Western New York Flash	5.080*
	(2.972)
Constant	3.080*
	(1.751)
Observations	42
R^2	0.689
Adjusted R^2	0.544
Residual Std. Error	2.402 (df = 28)
F Statistic	4.761*** (df = 13; 28)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Best Model for Wins

	Dependent variable:
	wins
total_alloc	1.539^{***}
	(0.349)
season	0.078
	(0.415)
Constant	−152.803
	(837.431)
Observations	42
R^2	0.345
Adjusted R^2	0.312
Residual Std. Error	2.951 (df = 39)
F Statistic	10.292^{***} (df = 2; 39)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Best Model for League Ranking

	Dependent variable:
	league_ranking
total_alloc	-1.189^{***}
	(0.275)
season	−0.022
	(0.309)
mean_min	−0.011
	(0.010)
mean_starts	0.967
	(0.880)
Constant	52.821
	(623.319)
Observations	42
R^2	0.415
Adjusted R^2	0.352
Residual Std. Error	2.155 (df = 37)
F Statistic	6.574^{***} (df = 4; 37)
Note:	*p<0.1; **p<0.05; ***p<0.01