Mathematics and Sports

An analysis of methods used to measure recruiting classes of major college football programs and assign star ratings to recruits

David M. McClendon[1,*] and Michael Nadrowski[2]

[1]Department of Mathematics, Ferris State University, ASC 2021, Big Rapids, MI 49307, [2]Department of Mathematics, Ferris State University, ASC 2021, Big Rapids, MI 49307

*Corresponding Author E-mail:
mcclend2@ferris.edu

Abstract

Previous studies have found correlation between the overall quality of a college football program's recruiting class and its performance on the field. These studies use as their measurement of a program's recruiting class either the total “points” accrued by the class, or the numbers of players in the class with various “star” ratings. In this paper, we study mechanisms used by 247Sports.com to produce these two measurements.

First, we examine variants of the formula used by 247Sports.com that estimates the quality of a recruiting class by computing a weighted sum of the ratings of its members. We find that using a weight function with smaller spread parameter, that takes much less information about the class into account, predicts team success for major college programs from 2016 to 2019 as well as, if not better than, the 247Sports.com approach.

Second, we search for the most effective thresholds for classifying players into “star”-like categories, finding that an optimal description of “blue chip” recruits is narrower than the one currently used. Furthermore, we find that dividing players into three different “star”-like classes does not produce a model that is substantively more predictive of team success than using only two such groups.

Keywords: ordinary least-squares regression, college football recruiting, star ratings

---

1 Introduction

College football teams across the United States compete not only on the field, but in recruiting—trying to convince the best high school players to join their program. The players who choose to attend a given school in a given year are called that school's recruiting class. It is commonly believed that the better a program's recruiting class is, the better its team will perform in subsequent years. As such, teams spend substantial resources (money and time) on scouting and recruiting players, and team recruiting activity is monitored closely by fans of the program.

247Sports.com (henceforth referred to as 247) is a website owned by CBS Sports that gathers and reports information on which schools players are interested in attending. They also rate prospective players’ abilities; in particular, they record for each recruit a “composite player rating”, which is an average of a numerical score assigned to that recruit from three services: ESPN, Rivals and 247. That average score is scaled so that it ranges from .7 and 1, with 1 being the highest rating.

Based on this composite player rating, 247 assigns to each recruit a “star rating”. 247 calls any player whose composite rating is at least .9830 a 5 (*five-star*) player; a player whose rating ranges from .8900 to .9829 is called a 4 player; a player whose rating ranges from .7970 to .8899 is a 3 □player, and lower-rated players are 2 □players. Four and five-star players are collectively called *blue chip* recruits; it is widely believed that signing blue chip recruits is key to success in college football[1].

247 takes these ratings of individual players and uses those ratings to create two measurements of the overall quality of a school's recruiting class. One, called *AVG*, is simply the average rating of each player in a school's recruiting class (multiplied by 100). The second, which 247 calls "Points" and which we refer to as *PTS*, is the default measurement used by 247 to rank recruiting classes on its website. *PTS* is a weighted sum of the rankings of the individuals comprising a school's recruiting class, and is computed via the following algorithm:

Suppose a recruiting class consists of N players, whose composite ratings are $P_1, P_2, \ldots, P_N$, where $P_1 \geq P_2 \geq \ldots \geq P_N$. Setting $P_n = .7$ for all $n > N$ finishes the construction of a decreasing sequence $\{P_n\}$ of numbers between .7 and 1 corresponding to each class, where only finitely many numbers in the sequence are greater than .7. Next, let $w(m,b,x)$ be the density function of a normal r.v. with mean m and standard deviation b, but normalized so that the function has maximum value 1. In other words,

$$w(m,b,x) = \exp\left(\frac{-(x-m)^2}{2b^2}\right).$$

247 uses $m = 1$ and a slightly different value of b each year, based on the total number of recruits signed by college programs in that year. However, no matter the year, b tends to be very close to 9. Using the function $w(x)$ to weight the recruit rankings P_1, P_2, …, produces *PTS*. Formally:

$$PTS = \sum_{x} w(1,b,x)100\,(P_x - .7).$$

Thus *PTS* is a single data point which attempts to measure the overall quality of a team's recruiting class. While previous studies (Langelett, 2003; Herda et al. 2009; Maxcy 2013; Mankin et al. 2021) have measured the correlation between *PTS* (or similar measurements) and a team's quality, none that we are aware of actually study how effective the mechanism producing *PTS* from individual player ratings is at combining those ratings into an overall measurement of a recruiting class' quality. In fact, while 247 touts its formula as "... ensuring that all commits contribute at least some value to the team's score without heavily rewarding teams that have several more commitments than others"[2], they provide no justification as to why the weight function w is chosen to be Gaussian, or why the mean parameter m is chosen to have value 1, or why it is important that all recruits add value to a class' overall rating. Similarly, while past work (Bergman and Logan 2016, Dronyk-Trosper and Stitzel 2017) investigates the impact of individual players to team success based on the number of players signed with each star rating, we are aware of no analysis as to whether or not the ranges of player rating chosen for each star rating are preferable to other ranges, in terms of producing a statistic that would best correlate with team success. (247 claims[3] that 5 players are those that are the equivalent of a first-round pick in the National Football League (NFL) draft, and that 4 □players are those that are likely to be drafted by an NFL team. However, past work (Wheeler 2018) indicates no major correlation between recruit ratings and NFL success.)

This paper is concerned with two questions: first, would different choices of parameters m and/or b in the weight function w produce an aggregate class score that is more strongly correlated with team performance on the field than *PTS*? Do the parameters that optimize this correlation support 247's position that every player needs to substantially contribute to the rating of a class? Second, would different quantitative thresholds for dividing players into categories analogous to 5 , 4 , etc. would produce counts of recruits in these ranges that better predict the success of the team? In other words, should more, or less, recruits be considered "blue chip" than are currently? Also, how many different categories (akin to star classifications) are needed to produce an effective method for predicting team success?

2 Literature review

This century, research into the effects of successful recruiting on team performance has become a topic of interest, due to both the increased media exposure and popularity of major college sports and a significant rise in the amount of publicly available data on both recruit quality and team performance.

[1] See, for example, the articles "This time, with emphasis: Of course recruiting rankings matter" by Chris Hummer, published by 247Sports.com on February 4, 2018, and "Blue-Chip Ratio 2020: The 15 teams who can win a national title" by Bud Elliott, published on 247Sports.com on June 11, 2020.
[2] See, for example, the Team Ranking Explanation at https://247sports.com/Season/2020-Football/CompositeTeamRankings/
[3] See https://247sports.com/college/georgia/LongFormArticle/Georgia-Bulldogs-Recruiting-Everything-you-need-to-know-about-247Sports-Rating-Process-143324123/

Langelett (2003) examined the effects of recruiting classes ranked in the top 10 according to Allen Wallace and Tom Lemming on a team's final ranking in polls conducted by the Associated Press and *USA Today*, finding evidence suggesting that improved recruiting does impact team performance in subsequent years. Langelett also found that schools that do well in a given year tend to have stronger recruiting classes in the following years.

Dumond, Lynch and Platania (2008) studied the supply and demand of college football recruits, focusing on the decision-making process utilized by recruits when selecting their school. Like Langelett, they found a positive relationship between recruit quality and on-field success.

Herda et al. (2009) studied the relationship between the total Rivals and Scout (Scout was purchased by 247 in 2017) points rankings of all Division I college football teams from 2002 to 2006 and the subsequent Sagarin ratings of those teams. They found that total points rankings from Rivals and Scout estimated up to 45% of the variance in a teams' Sagarin rating and the number of games that team wins. Later work of Maxcy (2013) and Mankin, Rivas and Jewell (2021) also found that recruiting class rankings had a significant and positive effect on team performance.

Bergman and Logan (2016) performed a similar analysis as Herda et al., but controlled for school-specific effects. They found that controlling for between-school heterogeneity lowered the effect of recruit quality, but that the remaining effect was still significant.

In a 2015 paper, Dronyk-Trosper and Stitzel (2017) used a two-stage least squares approach to study the effect of recruiting on team success. By including new control variables, they found evidence that the benefits of recruiting are driven by team-specific effects, reinforcing the findings of Bergman and Logan.

In all this work, however, the metric used for evaluating the strength of a class was either the aggregate rating of the class, or the star ratings of recruits in the class. We were unable to find prior work which actually examines the procedure by which 247 combines the ratings of individual players into an overall class rating, or the cutoffs used by 247 to partition recruits into the various star ratings. It is these mechanisms that we investigate.

3 Data

First, we define the set of *major* college football programs to be the set of those playing in the Atlantic Coast Conference (ACC), Big Ten, Big 12, Pac-12, or Southeastern Conference (SEC), together with Notre Dame, who competes as an independent. All of these 65 programs are members of the NCAA's Division 1 Football Bowl Subdivision (FBS), and collectively have won every College Football Playoff national championship contested. There is also a substantive difference in the quality of recruits signed by these programs and the recruits signed by other Division 1 programs (see, for example, Table 2 in (Bergman and Logan, 2016)).

We gathered both the composite rating and the star rating of each individual recruit in the classes of major college programs from 2016 to 2019 from 247Sports.com. We also recorded 247's weighted average *PTS*, and the number of recruits of each star rating in each class, for each major college program between 2016 and 2019.

We pulled the end-of-season Sagarin ratings for major college football teams from 2016 to 2019, which we denote as *SAG*, from Jeff Sagarin's web page (hosted on the USA Today website). These ratings were chosen as our primary method of assessing the quality of college football teams for a few reasons: first, Sagarin ratings include all Division 1 college football teams (not just the top 25 teams like traditional human polls like those conducted by the Associated Press and ESPN). Second, previous research (Herda, 2009) shows recruiting rankings to be most strongly correlated with Sagarin ratings (as opposed to winning percentage, number of wins, etc., which are greatly influenced by the difficulty of a team's schedule). That said, we also collected the number of games each major program won between 2016 and 2019 (denoted *WINS*) from ESPN's website, cross-checking against school media guides to ensure accuracy.

Summary statistics for this data are given in Table 1. For example, in 2019, the average composite player rating of a recruit who was a member of a major college team's recruiting class was .8752. We also give, in Figure 1, the empirical distribution function of composite player ratings for all major college football recruits from 2016-19, with the various star classifications indicated; a histogram for the same data is provided in Figure 2.

	Mean	Standard deviation
247 Composite player rating		
2016-2019	.8682	.0457
2019	.8752	.0415
2018	.8725	.0433
2017	.8633	.0473
2016	.8622	.0487
# of 5 recruits per class	0.469	1.099
# of 4 recruits per class	5.196	5.228
# of 3 recruits per class	16.277	5.700
# of 2 recruits per class	0.758	1.217
Overall class rating *PTS*		
2016-2019	217.121	40.809
2019	219.552	37.785
2018	221.304	39.280
2017	213.620	41.262
2016	211.989	43.429
Sagarin rating *SAG*		
2016-2019	77.288	10.853
2019	77.300	11.636
2018	77.363	9.396
2017	77.703	10.744
2016	76.786	11.708
Annual wins	7.269	2.968

Table 1: Summary statistics.
Data consists of all major college programs from 2016 to 2019.
Total observations of composite player ratings: 5902.
Annual observations of *PTS* and *SAG*: 65.

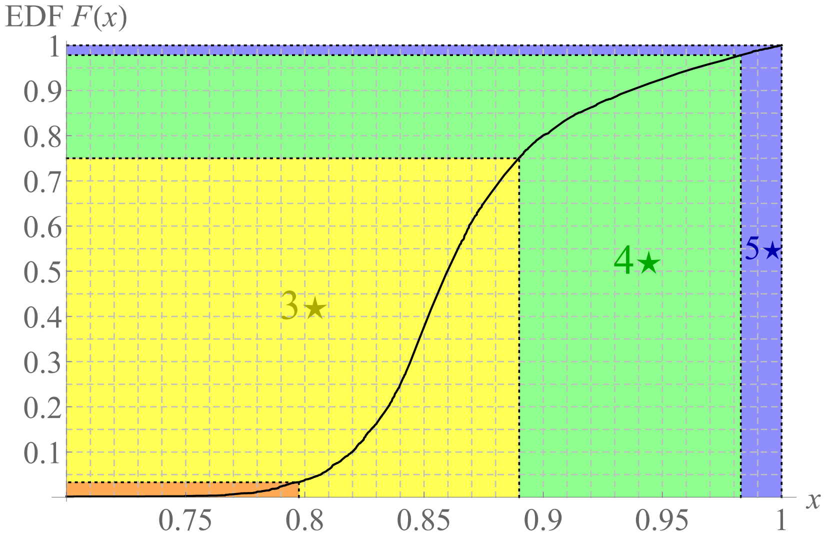

Figure 1: A graph of F, the empirical distribution function of composite player rating for players recruited by major college programs from 2016 to 2019. In particular,

$$F(x) = \frac{\#(players\ with\ composite\ rating\ \leq x)}{5902}$$

2 , 3 , 4 and 5 players are indicated in orange, yellow, green and blue, respectively.

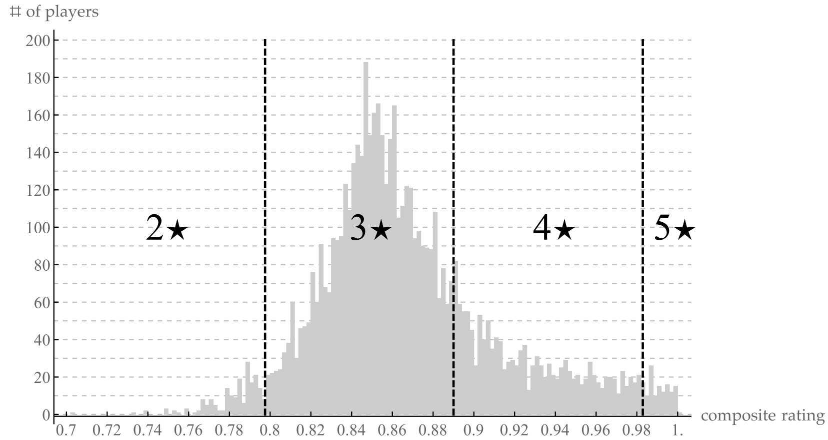

Figure 2: A histogram for composite player rating for players recruited by major college programs from 2016 to 2019.

4 Methodology

4.1 Computing the rating of a class from ratings of individuals

We first study three models (A, B and C) that use approaches based on the 247 scheme to compute an aggregate score for a team's recruiting class from the ratings of individual recruits.

4.1.1 Model A

In Model A, we seek the most predictive weight function *w* taken from the family with Gaussian curves whose maximum value occurs at $x = 1$. Toward that end, recall

$$w(1, b, x) = \exp\left(\frac{-(x-1)^2}{2b^2}\right).$$

Next, for a team whose recruiting class corresponds to the sequence $P_1, P_2, P_3, \ldots$ as described in Section 1, define

$$PTS_A(b) = 100 \sum_x w(1, b, x)(P_x - .7).$$

This $PTS_A(b)$ gives an analog of *PTS* for each team's recruiting class. Next, we performed an ordinary least-squares (OLS) regression between $PTS_A(b)$ and Sagarin rating *SAG* in the immediate following season, defining $R_A(b)$ to be the Pearson product moment correlation coefficient between $PTS_A(b)$ and *SAG*.

We used the computer algebra system *Mathematica* to numerically estimate the value b_A^* at which $R_A(b)$ is maximized: in particular, we ran the FindMaximum[ ] command with seed values that are multiples of .001, taking the largest value of $R_A(b)$ so obtained and calling this value R_A^*. We let b_A^* be the location of this maximum value and then computed the weighted average $PTS_A(b_A^*)$ for each class. We then used the R statistical software to compute OLS regressions between $PTS_A(b_A^*)$ and *SAG* and between $PTS_A(b_A^*)$ and *WINS*, testing parameters for statistical significance.

To provide some validation for our results, we also used 10-fold cross validation, repeated three times, to produce 30 different values of b_A^*, each coming from the procedure described above applied to 90% of the underlying data. For each of these values, we built the corresponding OLS model between $PTS_A(b_A^*)$ and *SAG* and then measured the total root mean squared error (RMSE) obtained when testing our model on the remaining 10% of the data. We also performed a data split, using the information from years 2016 to 2018 to find the optimal value b_A^* and then testing the corresponding OLS model on the 2019 data.

Last, we briefly investigated two other families of one-parameter weight functions (exponential and "half-tent"), using methodology similar to what is described above to find the value of the family's parameter the maximizes the correlation between *PTS* and *SAG*. In particular, to study the exponential family, set $w(\lambda, x) = e^{-\lambda x}$ and define

$$PTS(\lambda) = 100\sum_x w(\lambda, x)(P_x - .7).$$

As before, we used *Mathematica* to estimate the value of λ (denoted λ^*) for which the correlation between $PTS(\lambda)$ and SAG is maximized. For the half-tent family, we let

$$w(t, x) = \begin{cases} 1 - \frac{x}{t} & x \le t \\ 0 & \text{else} \end{cases},$$

defined $PTS(t)$ in the same way and located the value t^* where the correlation between $PTS(t)$ and SAG is maximized.

4.1.2 Model B

Here, our goal is to determine the best weight function w taken from the family of all Gaussian curves. So, following the notation from before, we set

$$w(m, b, x) = \exp\left(\frac{-(x-m)^2}{2b^2}\right)$$

and then define

$$PTS_B(m, b) = 100\sum_x w(m, b, x)(P_x - .7).$$

As before, we performed an OLS regression between $PTS_B(m,b)$ and SAG, defining $R_B(m,b)$ to be the correlation coefficient between $PTS_B(m,b)$ and SAG.

Using the FindMaximum[ ] command in *Mathematica* with seed values taken from the lattice of points whose x- and y-coordinates are multiples of .001, we numerically estimated values m_B^* and b_B^* which maximize $R_B(m,b)$ and denoted this maximum value as R_B^*. The weighted average $PTS_B(m_B^*, b_B^*)$ was computed for each recruiting class, and OLS regressions between $PTS_B(m_B^*, b_B^*)$ and SAG and between $PTS_B(m_B^*, b_B^*)$ and $WINS$ were performed, testing parameters for statistical significance. A 10-fold cross validation procedure akin to the one described in Model A was repeated in this setting, as was a data split between years 2016-18 and 2019.

Last, we investigated other families of weight functions (beta and "tent") to find the value of those families' parameters which maximize the correlation between PTS and SAG. To study the beta family, we used weight functions of the form

$$w(\alpha, \beta, x) = C\, x^{\alpha-1}(32 - x)^{\beta-1}$$

where C is a constant depending on α and β chosen so that the maximum value[4] of $w(\alpha, \beta, x)$ is 1. (We fixed the support of this function as the interval (0,32) because all recruiting classes studied had at most 31 members). For the tent family, we used piecewise-linear functions which pass through $(l,0)$, $(m,1)$ and $(r,0)$:

$$w(l, m, r, x) = \begin{cases} \frac{x-l}{m-l} & l \le x \le m \\ 1 - \frac{x-m}{r-m} & m \le x \le r \\ 0 & \text{else} \end{cases}.$$

We again performed the same kind of analysis, searching for (α^*, β^*) in the beta family and (l^*, m^*, r^*) in the tent family which maximize the correlation between PTS and SAG.

4.1.3 Model C

In this model, we adapt Model B by changing the floor of a recruit's ranking. Let $w(m,b,x)$ be as above, and let $f \in [.7,1]$. Next, we define

$$PTS_C(f, m, b) = 100\sum_x w(m, b, x)(\max\{P_x, f\} - f).$$

In this model, we take the ranking of any player whose rating is less than the "floor" f and reset it to f; then we compute the same weighted average as in Model B. As before, we perform an OLS regression between $PTS_C(f,m,b)$ and SAG to calculate the

[4] Of course, if α or β are less than 1, then $w(\alpha,\beta,x)$ has no maximum value. In practice, such choices of α and β do not lead to values of $R(PTS,SAG)$ which are close to optimal, so we ignore them.

correlation coefficient $R_C(f,m,b)$. Treating $R_C(f,m,b)$ as a function of m and b and as in Models A and B, we use *Mathematica* to estimate the maximum value $R_C^*(f)$ of $R_C(f,m,b)$.

4.2 Analysis of star categories

Here, we investigate the efficacy of various cutoffs used to partition recruits into "star" ratings.

4.2.1 Two-subset model

In this model, we divide all recruits into two classes, which will be labelled "5◆" and "4◆". To do this, we let $t_5 \in [.7,1]$ and for each recruiting class, let $N_5(t_5)$ be the number of players in that class whose composite rating is at least t_5. (We remark that since all individual player ratings are multiples of .0001, we need only consider values of t_5 that are multiples of .0001.) Then, let $N_4(t_5)$ be the number of players in each class whose composite rating is less than t_5. An OLS regression was performed with the model

$$SAG = \beta_5 \, N_5(t_5) + \beta_4 \, N_4(t_5) + \beta_0$$

and the correlation coefficient $R(t_5)$ for this model was computed as a function of t_5.

Setting t_5^* to be the value of t_5 at which this correlation coefficient is maximized, we decree a 5◆ recruit to be a player whose composite rating is at least t_5^*, and a 4◆ recruit to be a player whose composite rating is less than t_5^*. Least-squares regressions were then performed for *SAG* and *WINS* against $N_5(t_5^*)$ and $N_4(t_5^*)$, testing parameters for statistical significance.

We repeated this analysis using 10-fold cross validation with three repetitions, measuring the total RMSE in our OLS model. Last, we repeated this procedure with only the data from years 2016-18, using the 2019 data as a test set.

4.2.2 Three-subset model

Here, we partition recruits into three classes. Let $t_4, t_5 \in [.7,1]$ be such that $t_4 \le t_5$, and then set, for each class,

$$N_5(t_4, t_5) = \#(\text{recruits with composite rating in } [t_5, 1));$$

$$N_4(t_4, t_5) = \#(\text{recruits with composite rating in } [t_4, t_5));$$

$$N_3(t_4, t_5) = \#(\text{recruits with composite rating in } [.7, t_4)).$$

Then, we perform an OLS regression for the model

$$SAG = \beta_5 N_5(t_4, t_5) + \beta_4 N_4(t_4, t_5) + \beta_3 N_3(t_4, t_5) + \beta_0$$

(using both the entire data set and using 10-fold cross-validation with three repetitions) and compute the correlation coefficient $R(t_4, t_5)$ for this model as a function of t_4 and t_5.

Setting (t_4^*, t_5^*) to be the value of (t_4, t_5) at which this correlation coefficient is maximized, we define a 5 ♥ recruit to be a player whose composite rating is at least t_5^*, a 4 ♥ recruit to be a player whose composite rating is in $[t_4^*, t_5^*)$ and a 3 ♥ recruit to be a player with a composite rating less than t_4^*. Least-squares regressions were then performed for *SAG* and *WINS* against $N_5(t_4^*, t_5^*)$, $N_4(t_4^*, t_5^*)$ and $N_3(t_4^*, t_5^*)$, testing parameters for statistical significance.

5 Results

5.1 Model A

When considering all major programs from 2016 to 2019 as a single data set, the value of b^* yielding the strongest correlation between a Gaussian weighted sum of individual player rankings with mean parameter 1 and team Sagarin rating is $b_A^* = 6.882$. While this is quite a bit less than the parameter $b \approx 9$ used by 247, the corresponding correlation $R_A^* = .5871$ is only slightly larger than the correlation .5818 obtained when $b = 9$.

The values of b_A^* stay fairly consistent over the individual years studied, and the value of R_A^* increases somewhat from 2016 to 2019. These values of b_A^* for the entire sample are given in Table 2. However, when considering the sample conference-by-conference, the values of b_A^* and R_A^* vary greatly: Like the 247 model, Model A finds that the correlation between recruiting class rating and Sagarin rating is substantially higher for teams in the SEC than it is for teams in other conferences.

			Model A		Model B		
	N	$R(1,9)$	b_A*	R_A*	m_B*	b_B*	R_B*
All major programs							
2016-2019	260	.5818	6.882	.5871	8.038	1.752	.6006
2019	65	.6830	7.350	.6860	6.410	0.248	.7032
2018	65	.5775	6.667	.5833	8.571	2.410	.5940
2017	65	.5386	6.367	.5536	7.415	0.247	.5770
2016	65	.5449	7.509	.5469	11.016	0.185	.5763
ACC, 2016-2019	56	.4740	4.712	.5065	4.410	0.244	.5239
Big 12, 2016-2019	40	.5344	0.426	.5736	1.318	0.257	.5736
Big Ten, 2016-2019	56	.6164	7.032	.6221	7.011	0.184	.6415
Pac-12, 2016-2019	48	.4615	4.666	.4924	4.014	0.183	.5312
SEC, 2016-2019	56	.7483	9.645	.7488	10.273	0.243	.7814

Table 2: Approximate values of *m* and/or *b* in Models A and B which maximize the correlation between the weighted totals of recruiting classes and the Sagarin ratings of teams in that year.

Table 3 shows regression coefficients for OLS analysis of *SAG* and $PTS_A(b_A*)$, and between *WINS* and $PTS_A(b_A*)$. We estimate that a team's Sagarin rating increases by .1902 per unit of increase in $PTS_A(b_A*)$, and that a school wins .0440 more games per unit increase of $PTS_A(b_A*)$. These regression coefficients are larger than the ones obtained from the 247 model, and statistically significant at the .001 level. In particular, our study estimates that $PTS_A(b_A*)$ accounts for approximately 34.5% of the variance in a team's Sagarin rating (a slight increase from the 31.9% estimated using the 247 model) and 24.6% of the variance in the number of games a team wins (the 247 model estimates 23.1% of this variance). Furthermore, even when controlling for conference membership, the regression coefficients change only a small amount and remain statistically significant; see Table 4.

X	Y	β_1	N	R^2
PTS	*SAG*	0.1501*** (0.01367)	260	.3185
$PTS_A(b_A*)$	*SAG*	0.1902*** (0.01633)	260	.3446
$PTS_B(m_B*,b_B*)$	*SAG*	0.4250*** (0.03523)	260	.3607
PTS	*WINS*	0.0349*** (0.00397)	260	.2309
$PTS_A(b_A*)$	*WINS*	0.0440*** (0.00479)	260	.2463
$PTS_B(m_B*,b_B*)$	*WINS*	0.0986*** (0.01037)	260	.2595

Table 3: Regression results from six ordinary least-squares regression analyses, each fitting the equation $Y = \beta_0 + \beta_1 X$, where *X* and *Y* are as given in the first two columns of the chart. *PTS* refers to the 247 weighted total "Points"; $PTS_A(b_A*)$ and $PTS_B(m_B*,b_B*)$ are weighted totals using the ideal parameters found in Models A and B, respectively.
Note: Data is for all Power 5 conference teams for years 2016-2019. Standard errors are in parentheses.
*** $p < 0.001$

X	Y	β_1	β_{CONF}	N	R^2
PTS	SAG	0.1502*** (0.01425)	-.1073 (.41042)	255	.3149
$PTS_A(b_A*)$	SAG	0.1900*** (0.01695)	-.0668 (.40069)	255	.3362
$PTS_B(m_B*,b_B*)$	SAG	0.4231*** (0.03647)	-.0201 (.39475)	255	.3566
PTS	$WINS$	0.0353*** (0.00411)	-.0941 (.11840)	255	.2299
$PTS_A(b_A*)$	$WINS$	0.0443*** (0.00494)	-.0826 (.11669)	255	.2456
$PTS_B(m_B*,b_B*)$	$WINS$	0.0989*** (0.01066)	-.0719 (0.1154)	255	.2574

Table 4: Regression results from six ordinary least-squares regression analyses, each fitting the equation $Y = \beta_0 + \beta_1 X + \beta_{CONF} CONF$, where X and Y are as given in the first two columns of the chart. PTS, $PTS_A(b_A*)$ and $PTS_B(m_B*,b_B*)$ are as in Table 3; $CONF$ is a constant depending only on the team's conference.
Note: Data is for all Power 5 conference teams, not including Notre Dame, for years 2016-2019. Standard errors are in parentheses.
*** $p < 0.001$

5.1.1 Cross-validation in Model A

The 30 values of b_A* obtained from our 10-fold cross validation analysis averaged 6.891, very close to the b_A* obtained from the entire sample, and the standard deviation of these values of b_A* was .0356. The total root mean squared error (RMSE) was 11.302, which is close to the standard deviation of team Sagarin rating absent any other analysis.

5.1.2 Model A with split data

When using the data from 2016 to 2018 as a test set, we found that the correlation between $PTS_A(b_A)$ and SAG had a maximum value of $R_A* = .5572$, occurring when $b_A* = 6.662$. Using the corresponding OLS model to project the Sagarin rating of 2019 teams, we obtained a RMSE (root mean square error) in the estimate of 8.772, less than the standard deviation of 2019 team Sagarin rating without other analysis.

5.1.3 Other one-parameter families

For the exponential family of weight functions, we found the correlation between $PTS(\lambda)$ and SAG achieved its maximum value of .5776 when $\lambda* = .1723$. For the half-tent family, the maximum correlation of .5892 was achieved when $t* = 18.000$. Graphs of the optimal Gaussian, exponential and half-tent weight functions can be found in Figure 3 below. Observe that the optimal half-tent weight function is qualitatively similar to the optimal Gaussian weight function, but the optimal exponential function decreases more sharply.

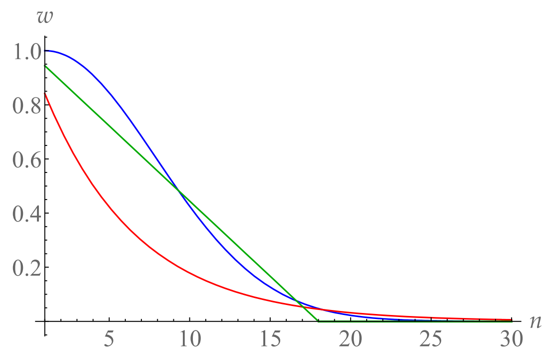

Figure 3: Graphs of optimal weight functions for our mean 1 Gaussian, exponential and half-tent models. The function $w(1,b*,x)$ is shown in blue, $w(\lambda*)$ is in red, and $w(t*)$ is in green.

5.2 Model B

When considering all major programs from 2016 to 2019 as a single data set, parameters $m^* = 8.038$ and $b^* = 1.752$ produce the strongest correlation between Gaussian weighted sum of individual player rankings and team Sagarin rating (see Table 2).

There is substantial variance in the values m_B^* and b_B^* over the individual years we studied, and the value of R_B^* is fairly consistent from 2016 to 2018 before a sharp increase in 2019. When considering schools only within individual conferences, however, the values of m_B^* and b_B^* vary greatly. As with the 247 model and Model A, when applying Model B we find that the correlation between recruiting class rating and Sagarin rating is substantially higher for teams in the SEC than it is for teams in other conferences.

Table 3 includes regression coefficients for OLS analysis of *SAG* and *WINS* against $PTS_B(m_B^*,b_B^*)$. We estimate that a team's Sagarin rating increases by .425 per unit of increase in $PTS_B(m_B^*,b_B^*)$, and that a school wins .0986 more games per unit increase of $PTS_B(m_B^*,b_B^*)$. These regression coefficients are larger than those obtained in Model A, and statistically significant at the .001 level. We find that $PTS_B(m_B^*,b_B^*)$ accounts for 36.1% of the variance in a team's Sagarin rating and 26.0% of the variance in the number of games that a team wins. As with Model A, we find little change in these regression coefficients when conference membership is controlled for (see Table 4), and these coefficients remain significant at the .001 level.

5.2.1 Cross-validation in Model B

In our 10-fold cross-validation approach to Model B, we produced 30 values of (m_B^*, b_B^*) whose average value was (7.713,1.213). The m_B^* so obtained had a standard deviation of .6606, and the b_B^* had a standard deviation of .9124. Of note, the average values of m_B^* and b_B^* obtained here are quite different than the values obtained from the entire sample. To explain why, we consider the behavior of the function $R_B(m,b)$. As reported earlier, the absolute maximum value of this function is .6006, occurring at $M_1 = (8.038,1.752)$. However, R_B also possesses a local maximum value of .6004 at the point $M_2 = (7.010,0.182)$, as shown in Figure 4 below. When performing cross-validation analysis, the values of (m_B^*, b_B^*) coming from each group of 9 folds clearly partition into two sets: 17 of the 30 values clustered around M_1 and the other 13 which are all close to M_2, as shown in Figure 5. The 17 points closest to M_1 have an average value of (8.234,1.960), which is relatively close to M_1; the coordinate-wise standard deviation of these 17 points is (.3107,.3661), substantially lower than the standard deviation of the entire sample. The other 13 points average (7.033,0.237) which is close to M_2, and have a standard deviation of (.2009,.1061) which is quite smaller than that of the entire sample. What seems to be happening here is that when one selects a subset of the entire sample, the corresponding function $R_B(m,b)$ generated by that subset has two local maxima, one near M_1 and one near M_2, which are very close in height. Which of these points is the location of the *global* maximum is sensitive to the subset of the data that is sampled.

The total RMSE from our 30 models was 8.802.

5.2.2 Model B with split data

When using the data from 2016 to 2018 as a test set, we found that the correlation between $PTS_B(b_B)$ and *SAG* had a maximum value of $R_B^* = .5749$, occurring at $(m_B^*, b_B^*) = (7.439,0.257)$ which is close to the point M_2 discussed in the previous section. Using the corresponding OLS model to project 2019 teams' Sagarin rating, we obtained a RMSE in the estimate of 9.352.

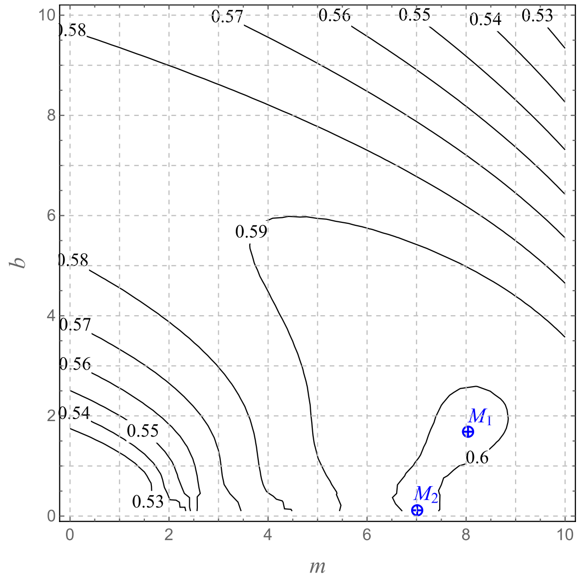

Figure 4: A contour plot for the function $R_B(m,b)$ that measures the correlation between $PTS_B(m,b)$ and SAG. $R_B(m,b)$ has a global maximum value of .6006 at the point $M_1 = (8.038,1.752)$ and a local maximum value of .6004 at $M_2 = (7.010,0.182)$.

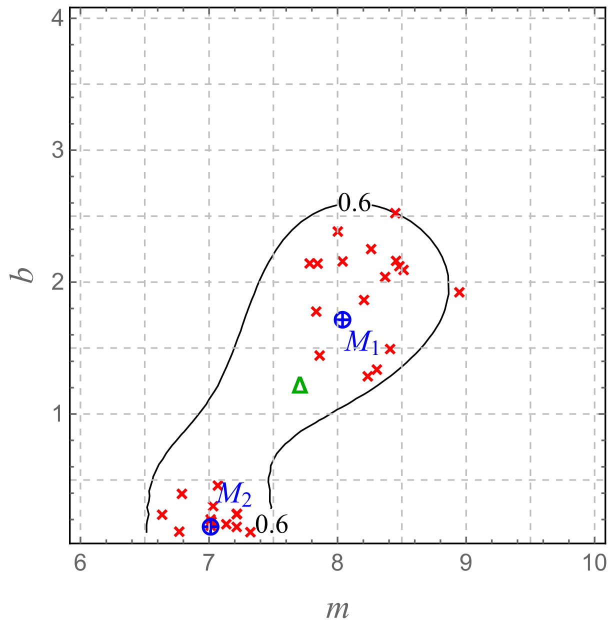

Figure 5: Each red X indicates one of the 30 values of (m_B*, b_B*) coming from our 10-fold cross-validation approach in Model B. The green triangle at (7.713,1.213) is the average value of those 30 points.

5.2.3 Other families

For the beta family, we found the maximum correlation between *PTS* and *SAG* was .6007, occurring when $\alpha^* = 17.211$ and $\beta^* = 51.6105$. For the tent family, the maximum correlation was .6012, occurring when $l^* = 5.280$, $m^* = 6.205$ and $r^* = 13.000$. Graphs of the optimal Gaussian, beta and tent weight functions can be found in Figure 6; note the qualitative similarities between the three graphs.

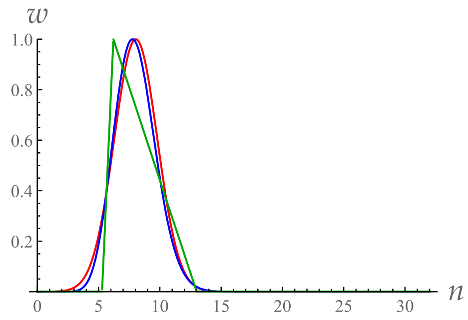

Figure 6: Graphs of optimal weight functions for our Gaussian (Model B), beta and tent models. The function $w(m^*,b^*,x)$ is shown in blue, $w(\alpha^*,\beta^*,x)$ is in red, and $w(l^*,m^*,r^*,x)$ is in green.

5.3 Model C

A graph of R_C^* is shown in Figure 7. Of note, when $f \leq .8$, the value of $R^*(f)$ is virtually unchanged. Also, the value of R_C^* remains at least 90% of $R_C^*(.7)$, until the floor f is raised to over .935, and the rate of decrease of R^* is quite small until f reaches approximately .94.

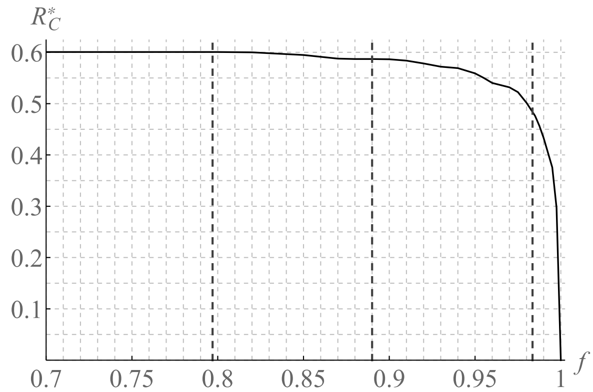

Figure 7: A graph of the function R_C^*, which gives the maximum correlation in Model C with the floor of a recruit's ranking set equal to f. The dashed lines indicate cutoffs used by 247 to classify players into star ratings.

5.4 Two-subset model

A graph of $R^2(t_5)$ against t_5 is given in Figure 8. The correlation between a team's Sagarin rating and its number of players in each group is maximized when $t_5^* = .9184$; the maximum value of $R^2(t_5)$ is $R^2(t_5^*) = .3600$. Using t_5^* as a threshold between 5◆ and 4◆ recruits, we see that 14.1% of all players recruited from 2016 to 2019 were 5◆ recruits; this is an average of 3.200 such players per class (Table 5 gives summary statistics for numbers of 5◆ and 4◆ recruits).

Regression results for *SAG* and *WINS* against the number of 5◆ and 4◆ recruits in a class are given in Table 6. We find that each additional 5◆ recruit increases a team's Sagarin rating by 1.172 on the average, and produces .249 additional wins for that team on the average; these regression coefficients are significant at the .001 level. Each additional 4◆ recruit, on the average, corresponds to a decrease in a team's Sagarin rating by .348 and in its number of games won by .113; these regression coefficients are significant at the .001 and .1 levels, respectively.

Of note, there exists negative correlation between the numbers of 5◆ and 4◆ recruits comprising a team's class (R(5◆, 4◆) = -.7619). This is as expected, since recruiting more 5◆ recruits leaves less room for 4◆ recruits in the same class. This correlation leads to a moderate variance inflation factor (VIF) in our model of 4.200. However, we can account any multicollinearity by performing an OLS regression between SAG (or WINS) and the number of 5◆ recruits (omitting the number of 4◆ recruits). The results are shown in Table 7; in this model we estimate that each 5◆ recruit increases a team's Sagarin rating by 1.455 and its number of games won by .3410. These regression coefficients are significant at the .001 level.

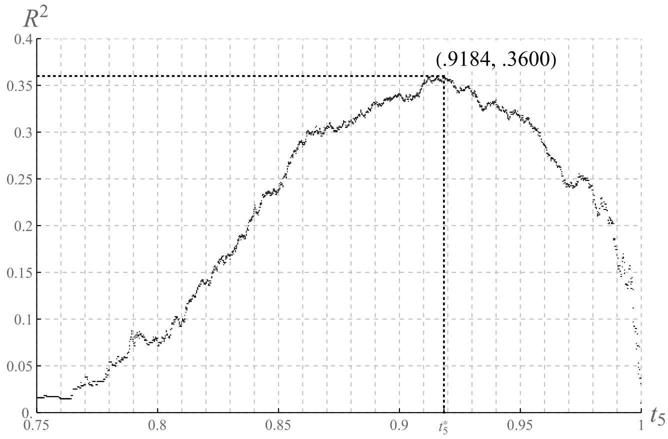

Figure 8: A graph of $R^2(t_5)$, the square of the correlation coefficient of the OLS estimate of a team's Sagarin rating from the number of recruits in a team's class whose rating is at least t_5 and the number of recruits in the class with rating less than t_5.

	Number of recruits	Percentage of all recruits	Average per class	Standard deviation per class
5◆ recruits	832	14.10%	3.200	4.388
4◆ recruits	5070	85.90%	19.500	4.825
5♥ recruits	830	14.06%	3.192	4.381
4♥ recruits	3590	60.83%	13.808	4.504
3♥ recruits	1482	25.11%	5.700	4.918

Table 5: Summary statistics for the classifications using our ideal two-subset and three-subset models.
Data consists of all major college programs from 2016 to 2019.
Total observations of composite player ratings: 5902.

Dependent variable	*SAG*	*WINS*
intercept	80.3248*** (3.9015)	8.6717*** (1.1375)
5◆ recruits (rating ∈ [.9184,1])	1.1718*** (0.1905)	0.2493*** (0.0555)
4◆ recruits (rating ∈ [.7, .9184))	-0.3480* (0.1733)	-0.1128* (0.0505)
N	260	260
R^2	.3601	.2726

Table 6: Regression results for team Sagarin rating and number of wins against recruit quality, as characterized by our ideal two-subset model. The VIF for both dependent variables is 4.200.
* $p < .1$; ** $p < .01$; *** $p < .001$.

Dependent variable	*SAG*	*WINS*
intercept	72.5812*** (0.6754)	6.1633*** (0.196)
5◆ recruits (rating ∈ [.9184,1])	1.4550*** (0.1248)	0.3410*** (0.03624)
N	260	260
R^2	.3484	.2585

Table 7: Regression results for team Sagarin rating and number of wins against recruit quality, as characterized by number of 5◆ recruits.
* $p < .1$; ** $p < .01$; *** $p < .001$.

As a point of comparison, we include regression results for *SAG* and *WINS* against the number of blue chip (4 and 5) recruits and number of non-blue chip recruits per class in Table 8. The regression coefficients corresponding to the number of recruits in the "top" group in our ideal two-subset model are substantively larger than those coming from the 247 classification. Also, the R^2 value for *SAG* in the ideal two-subset model is about 9% greater than the R^2 from the 247 approach.

Dependent variable	*SAG*	*WINS*
intercept	79.4728*** (3.9864)	8.4620*** (1.1557)
blue chip recruits (rating ∈ [.89,1])	0.7144*** (0.1810)	0.13952** (0.0525)
non-blue chip recruits (rating ∈ [.7,.89))	-0.3659* (0.1781)	-0.1164* (0.0516)
N	260	260
R^2	.3299	.2468

Table 8: Regression results for team Sagarin rating and number of wins against recruit quality, as characterized by whether or not the player is blue chip (where "blue chip" means having 247 star rating of 4 or 5).
* $p < .1$; ** $p < .01$; *** $p < .001$.

Our cross-validation analysis yielded similar data as that coming from the entire sample: the 30 values of t_5* obtained averaged .9173 and had standard deviation 0.002; the corresponding maximum values of $R^2(t_5^*)$ averaged .3614 with a standard deviation of .0157. Our values of β_5 (the regression coefficient for the number of 5◆ recruits) averaged 1.1787 and had standard deviation .1320, and our 30 values of β_4 (the regression coefficient for the number of 4◆ recruits) had mean -0.2880 with standard deviation 0.0673. The total RMSE for the Sagarin ratings estimated by our model was 9.814.

When using only the data from 2016 to 2018, we obtained $t_5^* = .9184$ (the same as for the entire sample) and $R^2(t_5^*) = .3487$, leading to the model

$$SAG = 84.8082 + 0.8777\ (\#\ 5◆\ \text{recruits}) - 0.5277\ (\#\ 4◆\ \text{recruits}).$$

Here, the intercept and the regression coefficient on the number of 5◆ recruits were significant at the .001 level, and the coefficient on the number of 4◆ recruits was significant at the .01 level. However, these regression parameters are quite different than those we obtained when looking at the entire data set or at the random folds in our cross-validation analysis, and this model produced a RMSE of 9.182 when tested on the 2019 data.

5.5 Three-subset model

A graph and contour plot of $R^2(t_4,t_5)$ can be seen in Figure 9. We found that the correlation between *SAG* and the number of players in each category is maximized when $t_4^* = .8401$ and $t_5^* = .9185$. Using t_4 and t_5 as thresholds between 3♥, 4♥ and 5♥ recruits, we see that 14.06% of all players recruited from 2016 to 2019 were 5♥ recruits (averaging 3.192 such players per class) and 60.83% of players recruited in that time frame were 4♥ recruits (averaging 13.808 such players per class). Summary statistics for these counts are available in Table 5.

Regression results for *SAG* and *WINS* against the number of 5♥, 4♥ and 3♥ recruits in a class are given in Table 9. We found that each additional 5♥ recruit increases a team's Sagarin rating by 1.064 on the average and produces .225 additional wins for that team on the average; these regression coefficients are significant at the .001 level. Each additional 3♥ recruit, on the average, decreases a team's Sagarin rating by .588 and its number of games won by .168; these regression coefficients are significant at the .01 and .1 levels, respectively. Note that the regression coefficients corresponding to the number of 4♥ recruits were not significant at the .1 level.

Similar to our two-subset model, there is negative correlation between the dependent variables which count 5♥, 4♥ and 3♥ recruits. In particular, R(5♥, 4♥) = -.2224, R(5♥, 3♥) = -.5437 and R(4♥, 3♥) = -.4783. This leads to the variance inflation factors ranging from 2 to 3 given in Table 9.

For comparison, in Table 10 we give regression analysis for *SAG* and *WINS* against counts of recruits in each star category. Of note, our ideal two-subset and three-subset models produce larger values for R^2 than the four-subset model (the subsets being 5 , 4 , 3 and 2) utilized by 247.

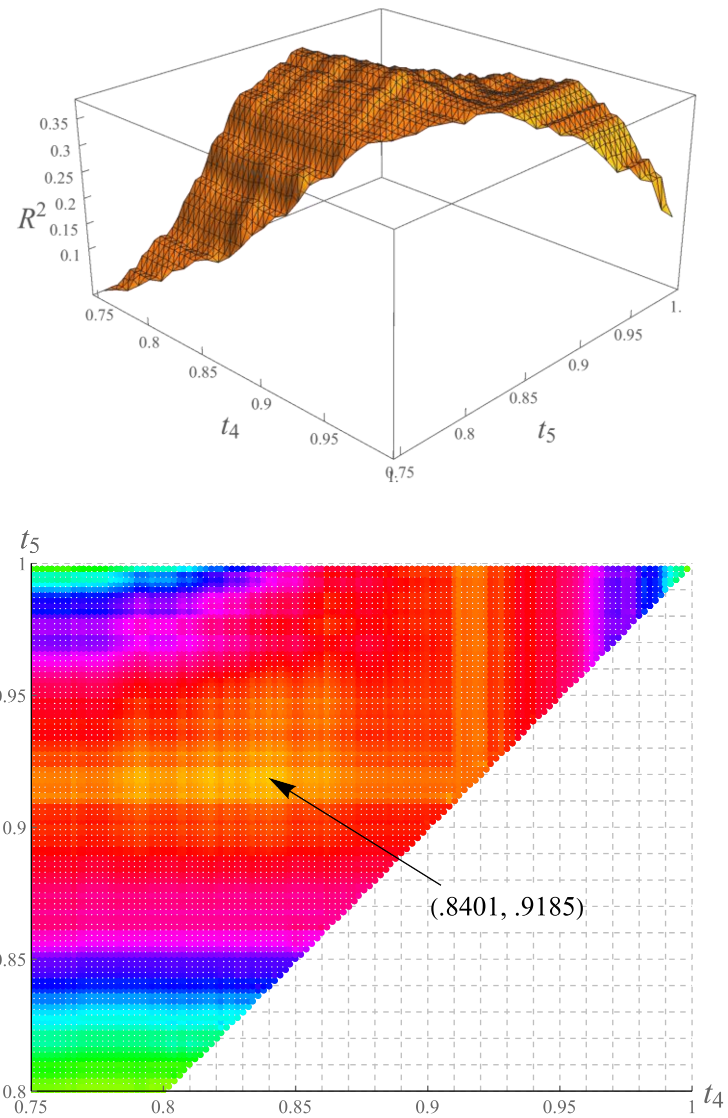

Figure 9: A graph and contour plot for the function $R^2(t_4, t_5)$ used in the three-subset model. The maximum value of R^2 occurs when the cutoff between 3♥ and 4♥ players is t_4^* = .8401, and the threshold between 4♥ and 5♥ players is t_5^* = .9185. This maximum value is R^2(.8401, .9185) = .3807.

Dependent variable	*SAG*	*WINS*
intercept	79.8616*** (3.8459)	8.5658*** (1.1289)
5♥ recruits (rating ∈ [.9185,1])	1.0640*** (0.1920) VIF = 2.479	0.2251*** (0.0564)
4♥ recruits (rating ∈ [.8401,.9185))	-0.1898 (0.1785) VIF = 2.271	-0.0771 (0.0524)
3♥ recruits (rating ∈ [.7,.8401))	-0.5877** (0.1899) VIF = 3.030	-0.1667* (0.0557)
N	260	260
R^2	.3809	.2867

Table 9: Regression results for team Sagarin rating and number of wins against recruit quality, as characterized by our ideal three-subset model. (VIF stands for

variance inflation factor.)
* $p < .1$; ** $p < .01$; *** $p < .001$.

Dependent variable	*SAG*	*WINS*
intercept	78.7860*** (3.9574)	8.3174*** (1.1550)
5 recruits (rating $\in [.983, 1]$)	2.1226** (0.6544)	0.4739* (0.1910)
4 recruits (rating $\in [.89, .983)$)	0.5358** (0.1943)	0.0988* (0.0567)
3 □recruits (rating $\in [.797, .89)$)	-0.2755 (0.1802)	-0.1009* (0.0526)
2 □recruits (rating $\in [.7, .797)$)	-1.0471* (0.4889)	-0.1869* (0.1427)
N	260	260
R^2	.3477	.2570

Table 10: Regression results for team Sagarin rating and number of wins against recruit quality, as characterized by 247 star rating.
* $p < .1$; ** $p < .01$; *** $p < .001$.

Our cross-validation analysis in the three-subset model yielded similar results as the entire sample. The 30 values of (t_4*, t_5*) obtained had mean (.8402, .9176) and standard deviation (.0002, .0018); the corresponding maximum values of R^2(t_4*, t_5*) averaged .3823 and had standard deviation .0171. Values of β_5 (β_j being the regression coefficient for the number of *j*♥ recruits in a class) had mean 1.0508 and standard deviation 0.0650; the 30 values of β_4 averaged -0.1958 with standard deviation 0.0654 and the values of β_3 had mean -0.5897 and standard deviation 0.0769. The total RMSE from the 30 folds studied was 8.7486.

When using only data from 2016 to 2018, we obtained (t_4*, t_5*) = (.8179, .9183) and R^2(t_4*, t_5*) = .3813, leading to the model

$$SAG = 84.6933 + 0.7759\ (\#\ 5♥\ \text{rec.}) - 0.4083\ (\#\ 4♥\ \text{rec.}) - 1.1519\ (\#\ 3♥\ \text{rec.}).$$

The intercept and the coefficients on counts of 5♥ and 3♥ recruits were significant at the .001 level; the coefficient on the count of 4♥ recruits was significant at the .1 level. Applying this model to the data from 2019 produced a total RMSE of 9.386.

6 Analysis

6.1 Models A, B and C

In Model A, we find that by changing only the spread parameter b, so long as $b \in [1, 20]$, the correlation between the corresponding weighted average $PTS_A(b)$ and *SAG* always remains between .52 and its maximum value R_A* = .5871. Furthermore, for $b \in [2.4, 12.5]$, $R_A(b)$ is more than 95% of R_A*. This suggests that when constructing weighted sums of individual player ratings using a Gaussian weight function with mean 1, the choice of spread parameter does not greatly influence the degree to which the weighted sum $PTS_A(b)$ correlates with team success.

Likewise, in Model B, we find that when using a Gaussian weight function with arbitrary mean and standard deviation to compute $PTS_B(m,b)$, we find that any choice of $(m, b) \in [0, 10] \times [.1, 10]$ leads to correlation values between .52 and the maximum value R_B* = .6006. This indicates that the choice of the parameters m and b in this procedure do not wildly affect how strongly $PTS_B(m,b)$ correlates with a team's Sagarin rating.

Furthermore, we found that optimal weight functions taken from other families produced values of R_A* close to those coming from Gaussian families, suggesting that there is nothing special about using Gaussian weight functions as opposed to weights taken from other families.

However, what stands out in Model B is that the optimal value of the spread parameter b_B* is much smaller than the corresponding parameter in either the 247 approach or Model A. This means that when using Model B, a much smaller amount of information (i.e. the ratings of a far smaller number of recruits) substantively contribute to the overall class rating. To measure this distribution more quantitatively, recall that

$$w(m, b, x) = \exp\left(\frac{-(x-m)^2}{2b^2}\right)$$

and set

$$F_{247}(n) = \frac{\sum_{x=1}^{n} w(1,9,x)}{\sum_{x=1}^{\infty} w(1,9,x)};\quad F_A(n) = \frac{\sum_{x=1}^{n} w(1,b_A^*,x)}{\sum_{x=1}^{\infty} w(1,b_A^*,x)};\quad F_B(n) = \frac{\sum_{x=1}^{n} w(m_B^*,b_B^*,x)}{\sum_{x=1}^{\infty} w(m_B^*,b_B^*,x)}.$$

The functions F_{247}, F_A and F_B therefore give the cumulative contribution to the total class rating coming from recruits who are among the n highest rated recruits in that class, according to the model used by 247, Model A and Model B, respectively. Graphs of F_{247}, F_A and F_B can be found in Figure 10.

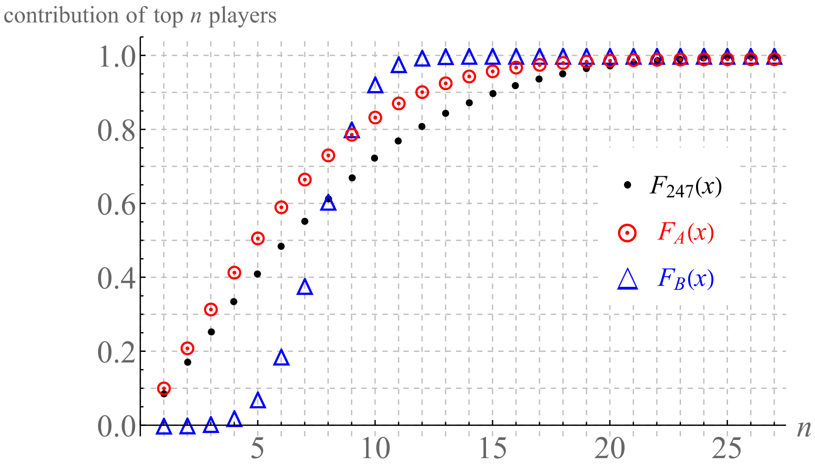

Figure 10: Cumulative contribution to overall class rating coming from a team's top n recruits, using the 247 weight function and our weight functions $w(1,b_A*,x)$ and $w(m_B*,b_B*,x)$ from Models A and B.

We observe that Model B gives substantial weight to far fewer players than the 247 model. Indeed, the values of P_n for $n \in \{5,6,\dots,12\}$ (i.e. the ratings of the 5th to the 12th best player in each class) contribute 97.5% of the overall class ranking when using the weight function $w(m_B*,b_B*,x)$ but only 47.6% of the overall class ranking using weight function $w(1,9,x)$ and only 48.8% of the overall class ranking in Model A. Furthermore, in Model B, players who are not among the 12 highest-rated recruits in a class collectively contribute only 0.5% of the overall class ranking. Yet, despite not taking these players into account very much, Model B is slightly more strongly correlated with team success than the 247 scheme. So it seems as though the assumption made by 247 that it is important to have all recruits contribute substantially to a team's overall ranking is incorrect, at least in terms of predicting team performance.

We conjecture three possible reasons for this. First, the rankings of the 5th to 12th best players in each class are likely to be strongly correlated with the rankings of the other players making up a recruiting class, so the ratings of other players provide essentially redundant information. Second, the evaluation process may take more time to judge the relative quality of the highest-rated recruits, and as such, these players' ratings may, in general, be more precise than those of lower-rated recruits. Third, players who make up the bottom half of a team's recruiting class may not be as likely to be significant contributors to a team (they may spend their career as reserves, backing up more talented recruits), so they have less impact on team success.

Our analysis of Model C similarly suggests that the ratings of top-rated players are substantially more important than the ratings of lesser-rated players in trying to predict team success from recruiting rankings alone. In particular, if we use a floor of $f = .8$ for a recruit's rating (rather than the .7 used by 247), the corresponding Gaussian weighted total would be almost equally correlated with team success as one that takes into account the particular ratings of recruits whose composite rating is less than .8. We believe this is for two reasons: first, few players (only 3.75% of the sample; see Figure 1) with a rating of less than .8 were signed by major programs from 2016 to 2019. Second, zero players with a rating of less than .8 were among the 12 best players in any class studied, meaning that players with a ranking of less than .8 barely contributed to the overall class rating produced by Model B.

However, Model C shows that even if we were to use a floor f such that many players in the top 12 of their class had a rating less than f the predictive value of the corresponding weighted total is still almost what it is when $f = .7$. For example, if we set $f = .89$ (the threshold used by 247 to distinguish 3 and 4 players), the corresponding value $R_C(.89) = .5870$ is still 97.7% of $R_B* = .6006$. Put another way, we find that without actually assigning individual ratings to any 3 or 2 players, one could obtain a model predicting team success that is 97.7% as effective as one that distinguishes such recruits.

Even further still, the value $R_C(f)$ only drops to less than 90% of its maximum value once $f \geq .96$. As only 5.61% of players in our sample have a rating of greater than .96, this means one could obtain a model predicting team success from only the ratings of the top 5.61% of recruits that is 90% as correlated with team success as a similar model accounting for the ratings of all recruits. This suggests that the efficacy of recruiting data in predicting team success comes mostly from observing what schools land a small subset comprising the most elite recruits.

6.2 Two-subset and three-subset models

We find that the choice of parameters (t_4 and t_5 here) often does not substantially impact the correlation between counts of recruits in each category and team success. For instance, in the two-subset model, we found that for $t_5 \in [.889, .934]$, the corresponding value of R^2 remained within 90% of its maximum value; this includes $t_5 = .89$ which is currently used by 247 to separate players into blue chip and non-blue chip categories. Nonetheless, our work suggests that recruits are best divided into two categories by using a higher rating, .9184, as the threshold between "blue chip" and "non-blue chip". This would place all 5 recruits and 67.1% of 4 recruits in the higher 5◆ category, and the remaining 32.9% of 4 players together with all 3 and 2 players in the lower 4◆ category. In particular, our analysis suggests that the top 2/3 of 4 players have a similar value to their team as a 5 recruit. This supports the findings of Bergman and Logan (2016), where the authors suggest that the impact of a 4 recruit is similar to that of a 5 recruit once fixed effects are accounted for.

Of note, we find in the three-subset model that our optimal value of t_5, .9185, is almost the same as the optimal value t_5* coming from the two-subset model (.9184). So our ideal three-subset model essentially partitions the 4◆ players from our ideal two-subset model into two groups: 4♥ and 3♥. Doing this generates a more predictive model, but only increases R^2 by about 5.78% from the R^2 from our two-subset model. So, mirroring what we find in Models A, B and C, we see that taking the time to more finely rate less regarded players is a task that produces only a small increase in predictive value. We suspect that considering a model with more subsets would continue to increase the corresponding R^2 value so obtained, but at a smaller and smaller rate, but did not pursue that analysis here.

6.3 Limitations on our study and further questions

6.3.1 Choices of metrics for recruit and team quality

In this paper we use players' composite ratings as independent variables. It is unclear as to the degree that said rating is actually a measure of the player's quality. For instance, scouts ostensibly evaluate players based on raw athletic measurements (40-yard dash times, shuttle times, bench press, etc.), by analysis of high school game film, by observing the player at camps, and even by investigating the player's academic accomplishments and character.[5] But it is not clear that different scouts have access to the same information or give these factors the same weight in their evaluation. Recruit ratings seem also not to be normalized across multiple years; as shown in our Table 1, the mean player rating of a major college recruit steadily increased from 2016 to 2019. Our work also does not account for "hidden gems" (Peng et al., 2018; Beckwith et al., 2019), that is, players who have lesser star ratings but have other attributes which may make them more likely to outperform peers with a similar rating.

Some also believe that players' ratings are inflated when they are offered and/or accept scholarships at programs having high profiles or that have a long history of fielding successful teams. One such hypothesis, called the "Bama bump" (Connelly, 2016), asserts that players who commit to Alabama receive a boost in their ratings. While to our knowledge this phenomenon has not been rigorously studied, if true it would raise the issue of whether ratings of recruits are independent estimates of a player's ability or simply a measurement of the degree to which that player is in demand by prominent college football programs.

Pro Football Focus (PFF) gives grades[6] to college players based on their performance during each play. The correlation between recruit rating and PFF grade is stronger for the highest-rated recruits (Beckwith et al., 2019), which makes sense given our findings that the highest-rated recruits are most strongly correlated with team quality. As such, it might be interesting to pursue how to predict team success using a weighted formula of the PFF grades of its members. However, PFF grades are not germane to our study, as we are interested in the relationship between ratings players receive *before* they play college football and their team's success (PFF grades do not exist for high school players).

In this paper we use Sagarin rating and number of wins as proxies for team quality (our dependent variable). Neither of these ratings account for the quality of coaching a player receives once he joins a college team, the facilities available to players, the budget of the team, etc. Nor do they account for injuries or lucky/unlucky breaks during games that are independent of whether the team is "good". We believe these issues are a major reason why none of our R^2 values exceed .4. While it is not the goal of our paper to produce a comprehensive model that most accurately predicts team success, it would be interesting to see how recruiting data based on our weighting schemes and/or classification into ◆ and/or ♥ categories could be incorporated into a more complex model trying to accurately predict team success, as in Bergman and Logan (2016) or Dronyk-Trosper and Stitzel (2017).

[5] See, for example, the article "A Scout's Take on How College Football Recruits are Evaluated" by Edwin Weathersby, published online by *Bleacher Report* on March 26, 2013.
[6] See http://pff.com/grades for a description of the methodology used by Pro Football Focus.

6.3.2 Possible synergies

We do not take into account possible synergies from combining top-level recruits at particular positions, nor are we taking into account the fact that a player's value may be different from one school to another. For instance, if Program X already has an outstanding quarterback, a highly-rated quarterback in the following class might not be as useful to Program X as he would be to Program Y that has no returning highly-rated quarterback.

Also, we do not factor in possible amplification that could take place when top-level recruits at particular positions are combined. Does, for example, the impact of a highly-rated quarterback increase when he is paired with a highly-rated wide receiver? We considered this question briefly during our study but found such synergies difficult to find, in part because programs that recruit well tended to recruit well at every position.

6.3.3 Other models

We limit ourselves to weighting schemes where the weighted player ratings are added. It might be interesting to consider formulas which multiply player ratings together (Tofallis 2014) or to use other weighting schemes for rank-ordered data (Barron and Barrett, 1996; Roszkowska 2013). Additionally, we consider only players' composite rating and do not separately consider their ratings from each of the three major services 247, ESPN and Rivals. It might be interesting to generalize our work by allowing the rating a player receives from each service as an independent variable.

6.3.4 Lagged effects

In our model, we only study the correlation between recruiting ratings and the Sagarin rating of a team in the immediate following season (i.e. the freshman year of the class studied). We ignore lagged effects, i.e. the impact a player recruited in year 2016 may or may not have on the Sagarin rating of his team in 2017 or 2018, and it is reasonable to ask if considering lagged effects might impact our analysis. However, Langelett (2003) found that the greatest impact of a recruiting class is in its freshman year (with discounted effects thereafter), and Bergman and Logan (2016) also found that lagged effects were small.

7 Conclusion

Our goal was to study the efficacy of two mechanisms in predicting college football team success: one used to convert the composite ratings of individual players into a metric evaluating the overall strength of a recruiting class, and one used to classify players into groups based on their individual rating.

As with previous studies, we find utility in the first mechanism, in that there exists positive correlation between Gaussian weighted sums of individual player ratings in a team's recruiting class and the performance of the team on the field. However, we find that these weighted sums are computed by a formula that takes into account an unnecessary amount of information. We argue that by using a Gaussian weight function with smaller spread parameter, and that therefore mostly takes into account the ratings of only a few players in the class, one can produce a weighted total that is more correlated with team success than one that that significantly takes into account the ratings of all the recruits in the class. We also found that other classes of weight function could be used as effectively as Gaussian densities, so long as they have similar qualitative characteristics. In addition, we find that considering only the ratings of a very small percentage of recruits produces a model that is almost as predictive as an analogous model that takes all player ratings into account.

Second, we find that the division of players into blue chip and non-blue chip categories would be a more effective predictor of team success if fewer players were categorized as "blue chip", and that further dividing non-blue chip players into various categories akin to star ratings is of only a small amount of additional value in predicting team success.

All told, our findings suggest that the composite rating of a high school football player can be a useful tool for determining the impact that player will have on his college team's success if that player is among the most elite players in the country, but is otherwise a much less useful data point.

Acknowledgements

The authors thank the referees for their helpful comments on the first draft of this article.

References

Barron, F. Hutton and Bruce E. Barrett. 1996. Decision quality using ranked attribute weights. *Management Science* **42**(11): 1515-1523. http://dx.doi.org/10.1287/mnsc.42.11.1515.

Beckwith, Gage, Tim Callahan, Bear Carlson, Tyler Fondren, R. Harris, Jacqueline Hoege, Tykai Martin et al. 2019. Systems analysis for University of Virginia football recruiting and performance. *2019 Systems and Information Engineering Design Symposium (SIEDS)*, 1-6. https://ieeexplore.ieee.org/document/8735611.

Bergman, Stephen A. and Trevon D. Logan. 2016. The effect of recruit quality on college football team performance. *Journal of Sports Economics* **17**(6): 578-600. https://doi.org/10.1177/1527002514538266.

Caro, Cary A. 2012. College Football Success: The Relationship between Recruiting and Winning. *International Journal of Sports Science & Coaching* **7**(1): 139-152. https://doi.org/10.1260/1747-9541.71.1.39.

Connolly, Bill. 2016. Does Alabama interest boost a recruit's rating? Of course, with good reason. *SBNation.com*. Retrieved from https://www.sbnation.com/college-football-recruiting/2015/2/3/7972129/alabama-recruiting-bama-bump-ratings-rankings on June 12, 2021.

Dumond, J. Michael, Allen K. Lynch and Jennifer Platania. 2018. An economic model of the college football recruiting process. *Journal of Sports Economics* **9**(1): 67-87. https://doi.org/10.1177/1527002506298125.

Dronyk-Trosper, Trey and Brandli Stitzel. 2017. Lock-in and team effects: recruiting and success in college football athletics. *Journal of Sports Economics* **18**(4): 376-387. https://doi.org/10.1177/1527002515577885.

Herda, Trent J., Eric D. Ryan, Jason M. DeFreitas, Pablo B. Costa, Ashley A. Walter, Katherine M. Hoge, Joseph P. Weir and Joel T. Cramer. 2009. Can recruiting rankings predict the success of NCAA Division I football teams? An examination of the relationships among Rivals and Scouts recruiting rankings and Jeff Sagarin end-of-season ratings in collegiate football. *Journal of Quantitative Analysis in Sports* **5**(4), 240-345. https://doi.org/10.2202/1559-0410.1200.

Langelett, George. 2003. The relationship between recruiting and team performance in division 1A college football. *Journal of Sports Economics* **4**(3): 240-245. https://doi.org/10.1177/1527002503253478.

Mankin, Jeffrey A., Julio A. Rivas and Jeffrey J. Jewell. 2021. The effectiveness of college football recruiting rankings in predicting team success: a longitudinal study. *Research in Business and Economics Journal* **14**, 4-22. Available at SSRN: https://ssrn.com/abstract=3187688.

Maxcy, Joel G. 2013. Efficiency and managerial performance in FBS college football: to the employment and succession decisions, which matters most, coaching or recruiting? *Journal of Sports Economics* **14**(4): 368-388. https://doi.org/10.1177/1527002513497170.

Peng, Kunrui, Jonathan Cooke, Annie Crockett, Dongmin Shin, Austin Foster, Jacob Rue, Roby Williams et al. 2018. Predictive Analytics for University of Virginia Football Recruiting. 2017 Systems and Information Engineering Design Symposium, 1-6. http://dx.doi.org/10.1109/SIEDS.2018.8374745.

Roszkowska, Ewa. 2013. Rank ordering criteria weighting methods – a comparative overview. *Studia Ekonomiczne* **5**(65): 14-33. http://dx.doi.org/10.15290/ose.2013.05.65.02.

Tofallis, Chris. 2014. Add or Multiply? A tutorial on ranking and choosing with multiple criteria. *INFORMS Transactions on Education* **14**(3): 109-119. https://doi.org/10.1287/ited.2013.0124.

Wheeler, Nicholas. 2018. Do high school football recruit ratings accurately predict NFL success? *CMC Senior Theses* **1846**, available at http://scholarship.claremont.edu/cmc_theses/1846.