1-1-2019

College Football: Doing Less With More and More With Less

Gary N. Smith
Pomona College

Jordan Hawkins
Pomona College

Jack Storrs
Pomona College

College Football: Doing Less With More and More With Less

Jordan Hawkins, Pomona College
Gary Smith, Pomona College
Jack Storrs, Pomona College

Corresponding author:
Gary Smith
Department of Economics
Pomona College
425 N. College Avenue
Claremont CA 91711
gsmith@pomona.edu

College Football: Doing Less With More and More With Less

Abstract

There is a substantial and highly significant correlation between the performance of widely followed college football teams and the pre-college recruiting scores received by their players. This correlation implies a regression toward the mean that should be taken into account in the identification of under-performing and over-performing teams and can also be used to improve pre-season predictions of the performance of teams with highly rated and lowly rated recruits.

College Football: Doing Less With More and More With Less

One of the most anticipated sporting events of the year is not an actual game, but college football's National Signing Day, where high school seniors commit to college football programs in return for athletic scholarships. Drama builds as recruited athletes announce their decisions throughout the day, sometimes surprising fans and coaches who are disappointed by lost prospects or elated by signed prospects. The most highly coveted players hold nationally broadcast press conferences. Is the hype and hysteria warranted? How well do recruiting successes and failures predict the performance of college football teams?

Background

Four major scouting services, Scout, Rivals, 247 and ESPN, grade college football prospects by star values ranging from 5-star (an immediate impact player, among the top 2 to 3 dozen prospects in the country) to 2-star (potential role player). Unknown players are rated 1-star.

Caro (2012) looked at how well recruiting success, as measured by a team's average star score, predicted victories in conference games. She found that differences in average star rankings explained 80, 63, and 78 percent of the variance in conference winning percentages within the Southeastern Conference (SEC), Big 10, and Big 12 conferences, respectively. Average star rankings were less successful in predicting performance in other conferences and did not produce statistically significant results.

Bergman and Logan (2013) studied the impact of top recruits on the performance of teams that are in the Division I Football Bowl Subdivision (FBS) of the National Collegiate Athletic Association (NCAA). Looking at data for the years 2002 through 2012, they found a substantial and statistically significant relationship between recruit quality and both field success and the chances of appearing in a post-season bowl game. They estimated that a 5-star recruit is worth more than $150,000 in expected bowl revenue.

Hummer (2013) identified several players who had been underrated when they were recruited. For example, Johnny Manziel was only a 3-star recruit but won the Heisman trophy as a college freshman and was a first-round pick in the National Football League (NFL) draft. Overall, however, Hummer found that Rival's recruiting rankings were useful in predicting team performance.

Meers (2013) addressed the perception that college football recruiting has a lagged effect on a team's success. He examined the 2002 through 2011 Rivals recruiting rankings and the 2006 through 2012 college football F/+ rankings. F/+ is a statistical measure of a team's success rate and efficiency invented by Bill Connelly and Brian Fremeau that takes into account the strength of the opponent. The offset time periods were used to account for the recruiting class of 2002 becoming seniors in the 2006 season and so on. Meers found that the sophomore and junior class recruiting rankings had a mild, and not quite statistically significant, effect on a team's performance.

Regression Toward the Mean

In educational testing (Lord and Novick 1968, Smith and Smith 2005), a student's ability μ is the statistical expected value of his or her test score, and a student's actual score X on any particular test differs from ability by an independent and identically distributed error term ε:

$$X = \mu + \varepsilon \tag{1}$$

If the error scores are independent of abilities, then the variance of the observed scores across students is equal to the variance of abilities plus the variance of the error scores:

$$\sigma_X^2 = \sigma_\mu^2 + \sigma_\varepsilon^2$$

Because the variance of scores is larger than the variance of abilities, observed differences in scores typically overstate the differences in abilities.

If we knew the students' abilities, Equation 1 could be used to make unbiased predictions of each student's score. However, teachers are interested in the reverse question, using observed scores to estimate unobserved abilities. Kelley's equation (Kelley, 1947) says that a student's ability can be estimated from a weighted average of the student's score and the mean score,

$$\hat{\mu} = \rho^2 X + \left(1 - \rho^2\right)\bar{X} \tag{2}$$

where the weight ρ^2 (the test's *reliability*) is the squared correlation between scores and abilities, which equals the ratio of the variance of abilities to the variance of scores:

$$\rho^2 = \frac{\sigma_\mu^2}{\sigma_X^2}$$

If a test's reliability were 1, so that scores and ability are perfectly correlated, each student's estimated ability would be equal to the student's test score. If a test's reliability were zero, so that scores are unrelated to ability, every student's estimated ability would equal the average test score because there is no way to distinguish between above-average and below-average students. For cases between these extremes, a student's estimated ability regresses toward the average test score, and there is more regression for less reliable tests.

In college football, those athletes with the highest and lowest recruiting scores are, on average, closer to the mean ability than their scores indicate. How much closer depends on the correlation between recruiting scores and true ability. If the correlation between recruiting scores and ability were 0.71, the squared correlation would be 0.5, and an athlete's estimated ability would be halfway between his recruiting score and the average recruiting score.

There is regression toward the mean whenever there is an imperfect correlation between predicted and actual events (Smith 2016). For example, because of the imperfect correlation between predicted and actual changes in interest rates, interest rate changes tend to be smaller than predicted (Dorsey-Palmateer and Smith 2007). Because of the imperfect correlation between predicted and actual corporate earnings, actual earnings tend to be closer to the mean than are predicted earnings (Keil, Smith, and Smith 2004).

In college football, the imperfect correlation between college football recruiting and college football performance implies that the teams that are most successful and least successful in recruiting tend to be closer to the mean in performance.

If we use the linear regression model to predict a team's performance *Y* from its recruiting score *X*,

$$Y = \alpha + \beta X + \varepsilon$$

The least squares estimates are

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{s_y}{s_x} r$$

where *r* is the correlation between *X* and *Y*. Therefore, the least-squares predictor of *Y*

$$\hat{y} = a + bx$$

can be rewritten in terms of standardized variables

$$\left(\frac{\hat{y}-\overline{y}}{s_y}\right)=r\left(\frac{x-\overline{x}}{s_x}\right) \tag{3}$$
$$\hat{z}_y = r\hat{z}_x$$

If the correlation between team recruiting scores and team performances is 0.7, then a team with a recruiting score one standard deviation above the mean is predicted to perform 0.7 standard deviations above the mean.

A Model

Players are generally eligible for four years of participation in intercollegiate sports, although injured players can apply for additional time. Some players delay their four years of athletic eligibility by "redshirting" during their first year in college; the *redshirt* label comes from the fact that non-roster players often wear red shirts when practicing with the team. Before 2018, redshirted football players could not play in any games during their redshirt year. Now, they are allowed to appear in up to four games without losing a year of eligibility.

A "true freshman" plays the first year; a "redshirted freshman" is a sophomore who was redshirted freshman year. A fifth-year senior is in the fifth year at college, but the fourth year playing sports.

Because redshirting is common in college football, we modeled the recruiting assets A_t of a team in year t as a function of the five most recent years of recruiting scores R_{t-i}, $i = 0$ to 4. Football is a fall sport, so the most recent recruits graduated from high school in the same calendar year as the current football season; for example, first-year college students in the fall of 2018 committed in 2018 and graduated from high school in 2018 (or earlier).

We assumed that the value of a team's recruiting assets in year t is an exponential moving average of the five most recent recruiting scores,

$$X_t = \sum_{i=0}^{4} \lambda^i R_{t-i} \tag{4}$$

where the parameter λ determines the exponentially decreasing weights.

We used a nonlinear maximum likelihood procedure to estimate λ as well as the other parameters of the linear regression model

$$\begin{aligned} Y_t &= \alpha + \beta X_t + \varepsilon_t \\ &= \alpha + \beta \sum_{i=0}^{4} \lambda^i R + \varepsilon_t \end{aligned} \tag{5}$$

where Y_t is a measure of the college team's success in year *t*.

Data

We used the 247 database to assess a team's recruiting success because 247 has more years of robust data than other databases such as Rivals or Scout. Based on the number of stars assigned to each recruit, 247 gives each college team a composite recruiting score for each recruiting class. A team's performance each year was gauged by Jeff Sagarin's Composite Ranking, a proprietary system which takes into account game scores, quality of opponents, and how long ago the games were played.

We excluded recruiting years before 2002, because these years only had unique scores for the top-50 schools with all other programs arbitrarily given the same recruiting score. We also excluded teams with incomplete recruiting or Sagarin data.

Table 1 summarizes our data, starting in 2006 because this is the first year that has five years of recruiting data. The number of teams covered has increased over time and the means and standard deviations have varied considerably from year to year. The standard deviations for both the Sagarin rankings and recruiting scores have trended upward over time as the top-tier teams have increasingly distanced themselves from lesser teams. We consequently calculated annual Z-scores for both the Sagarin and recruiting data using the respective means and standard deviations for each year.

Results

Figure 1 shows the standard error of estimate (SEE) for Equation (5) for values of λ ranging from 0.01 to 10. A value $\lambda = 1$ corresponds to weighting all 5 recruiting classes equally. Recent classes are given more weight if λ is less than 1; distant classes are given more weight if λ is larger than one. Specifically, $\lambda = 0.10$ corresponds to a recent year's recruiting ranking being weighted 10 times the previous year's ranking, while $\lambda = 10$ corresponds a recent ranking being weighted 1/10 the previous year's ranking.

For our data, the minimum SEE is at $\lambda = 0.49$, so that a recent year is weighted roughly twice as heavily as the previous year. Figure 1 shows that the SEE is not very sensitive to λ, perhaps because recruiting results do not vary substantially from one year to the next. In the extreme case where teams have the same success year after year, it wouldn't matter which years are weighted more heavily than others. Here, the SEE is only 1.5 percent lower with $\lambda = 0.49$ than with $\lambda = 1.00$, which corresponds to weighting all five recruiting classes equally, so we used an equal weighting which is easier for most fans to understand.

Figure 2 shows a scatter plot of Sagarin ratings and recruiting assets. Interestingly, there is more performance variation for teams with lower recruiting assets. This makes sense because low-rated teams are comprised mostly of largely unknown players.

The correlation in Figure 2 between recruiting and performance is 0.75, with a t-value of

44.08. Although the correlation is positive, as expected, it is substantially less than 1 due to a variety of factors, including errors in predicting the success of high school players when they play college football, the consequences of injuries and other life events, and the influence of coaching on player performance and team success.

As noted earlier, the magnitude of the regression toward the mean depends on the correlation between predicted and actual performance. Here, the 0.75 correlation implies that teams that are one standard deviation above or below the mean in recruiting tend to be only 0.75 standard deviations from the mean in performance.

Under-performing and Over-performing Teams

We should not gauge a team's under-performance or over-performance by a simple comparison of its performance score with its recruiting score; for example, expecting a team with a recruiting score two standard deviations above the mean to perform two standard deviations above the mean. Regression towards the mean tells us that the top and bottom recruiting teams are likely to perform closer to the mean. The relevant question is how well they performed relative to this anticipated regression. Specifically, the 0.75 correlation between performance and recruiting means that we should expect a team with a recruiting score 2 standard deviations above the mean to perform 1.46 standard deviations above the mean, and this is the appropriate benchmark for seeing whether a team is under-performing or over-performing.

Equivalently, we can use the vertical distance between a team's actual performance and its predicted performance shown by the fitted line in Figure 2 to gauge under-performance and over-performance. Figure 2 shows the fitted line and 95 percent prediction interval. Among those performances outside a 95 percent prediction interval, Table 2 shows five teams that deservedly are known as over-performers or under-performers.

Boise State’s 2010 team was rated 0.608 standard deviations below the mean in recruiting, but performed 1.810 standard deviations above the mean, winning 12 games and losing only 1. Boise State was coached by Chris Petersen for eight years (2005–2013) with an overall record of 92–12, despite only one season with an above-average recruiting class, a 0.09 Z score in his final year. Petersen was the first two-time winner of the Paul “Bear” Bryant Coach of the Year Award and a clear example of how some coaches can do more with less.

The 2010 Air Force team had a recruiting score that was rated 2.035 standard deviations below the mean and, taking regression into account, was predicted to perform 1.527 standard deviations below the mean, but Sagarin rated its performance (an 8-4 regular season record and an Independence Bowl win over Georgia Tech) as 0.553 standard deviations above the mean.

At the other end of the over-performance/under-performance spectrum, the 2008 Michigan team, 2012 Colorado team, and 2008 Washington team all did less with more.

Michigan began the 2008 season with the 10th ranked recruiting assets and a new head coach, Rich Rodriguez, who put in a spread option offense that didn’t work well with the players Michigan had recruited before his arrival. The team won 3 games and lost 9, the most single-season losses in Michigan’s 129-year history, and Sagarin ranked them 84th in performance. The team went from 1.725 standard deviations above the mean in recruiting to 0.652 standard deviations below the mean in performance.

The 2012 Colorado team’s recruits were judged 0.390 standard deviations above the mean, yet the team went 1–11, Colorado’s worst season ever, and the head coach Jon Embree was fired at the end of the season. The 2008 Washington team did even worse, with a recruiting score

0.503 standard deviations above the mean, and a performance that was 1.351 standard deviations below the mean. Washington was outscored 463 to 159 and lost every game, earning the dubious distinction of being the second winless team in the history of the PAC-10 league. The coach was fired.

Predicting Success

We used Equation 5 at the beginning of each season to predict a team's performance, taking the anticipated regression into account. Specifically, we predicted the performance scores each season based on the preseason recruiting scores and the previous season's correlation between performance scores and recruiting scores, information that is available to fans before the start of each season.

Predictive accuracy was gauged by the mean absolute error (MAE)

$$\text{MAE} = \frac{\sum_{t=1}^{n} \left| \hat{Y} - Y \right|}{n}$$

and the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} \left(\hat{Y} - Y \right)^2}{n}}$$

The regression principal does not change the rank order of predicted performance. The top recruiting team will not be predicted to finish tenth in performance; it will just be predicted to be closer to the mean in performance then it was in recruiting.

Table 3 separates the ratings for our 13 years of data into quintiles based on the annual recruiting scores. As might be expected, there is relatively little regression or forecasting improvement for the middle three quintiles. At the outer quintiles, however, there is substantial regression to the mean and room for improvement in predicting performance.

The teams in the top recruiting quintile are, on average, 1.380 standard deviations above the mean in recruiting, but only 1.038 standard deviations above the mean in performance. This $\frac{1.038}{1.380} = 0.75$ regression is consistent with the 0.75 regression predicted from the correlation between performance and recruiting. The mean absolute error is 17 percent higher if the recruiting score is used to predict performance without taking regression into account. The root mean square error is 18 percent higher.

Similarly, the teams in the bottom recruiting quintile are, on average, 1.251 standard deviations below the mean in recruiting, but their performance is only 0.893 standard deviations below the mean. This $\frac{0.893}{1.251} = 0.71$ 0 regression is again consistent with the 0.75 regression predicted from the correlation between performance and recruiting. The mean absolute error is 10 percent higher and the root mean square error is 13 percent higher if regression is neglected and the recruiting score alone is used to predict performance.

Conclusion

For widely followed college football teams, there is a 0.75 correlation between a team's performance and its recruiting assets, using a simple equal weighting of recruiting scores for the five most recent recruiting classes. This correlation is highly statistically significant and implies a substantial regression toward the mean, in that teams that are one standard deviation from the mean in recruiting are, on average, 0.75 standard deviations from the mean in performance. Assessments of under-performing and over-performing teams should be based on this anticipated regression. In addition, pre-season predictions of the performance of teams with highly rated and lowly rated recruits can be improved substantially by taking regression into account.

References

Anonymous, 2013, "Analyzing the Predictive Capability of Recruiting Rankings." *Sports Illustrated*, February 4, https://www.si.com/football/2013/02/04/recruiting-rankings-predictive-accuracy.

Bergman, Stephen A. and Trevon D. Logan. 2014. "The Effect of Recruit Quality on College Football Team Performance." *The Ohio State University*.

Caro, Cary A. 2012. "College Football Success: The Relation Between Recruiting and Winning." *International Journal of Sports Science & Coaching*, 7:1, 139-152.

Dorsey-Palmateer, Reid, Smith, Gary. 2007. "Shrunken Interest Rate Forecasts are Better Forecasts," *Applied Financial Economics*, 17 (6), 425-430.

Hummer, Chris. 2018. "How Recruiting Rankings Fare Projecting Future NFL Draft Picks." *247* Sports, April 30, https://247sports.com/Article/NFL-Draft-recruiting-rankings-go-hand-in-hand--117819292/

Keil, Manfred, Smith, Gary, Smith, Margaret H, 2004. "Shrunken Earnings Predictions are Better Predictions," *Applied Financial Economics*, 14 (13), 937-943.

Kelley, T. L. 1947. *Fundamentals of Statistics.* Cambridge, MA: Harvard University.

Lord, F. M., & Novick, M. R. 1968. Statistical theory of mental test scores. Reading, MA: Addison-Wesley.

Meers, Kevin. 2013. "Analyzing the Lagged Effect of College Football Recruiting," *Harvard Sports Analysis Collective*, February 21, https://harvardsportsanalysis.wordpress.com/2013/02/21/analyzing-the-lagged-effect-of-college-football-recruiting/

Smith, Gary. 2016. *What the Luck?: The Surprising Role of Chance in our Everyday Lives*. New

York: Overlook.

Smith, Gary, Smith, Joanna. 2005. “Regression to the Mean in Average Test Scores,” *Educational Assessment*, 10 (4), 377-399.

Table 1 Data Summary

		Sagarin Scores		Recruiting Scores	
Year	Teams	Mean	Std. Dev.	Mean	Std. Dev.
2006	111	71.31	11.96	147.13	66.52
2007	115	70.98	11.69	156.31	62.98
2008	116	71.38	10.88	164.30	61.56
2009	115	71.57	10.72	176.15	58.54
2010	115	71.30	12.01	184.52	52.72
2011	116	71.37	11.89	152.30	72.43
2012	118	71.10	11.39	171.09	63.07
2013	126	68.80	13.98	162.95	61.98
2014	135	68.18	14.06	156.38	67.27
2015	149	66.08	15.24	149.98	72.08
2016	185	62.13	16.98	130.87	76.16
2017	209	59.89	17.74	119.43	79.74
2018	217	59.38	17.29	119.05	86.04

Table 2 Over-performing/Under-performing Teams

Year	Team	Recruiting Assets	Predicted Sagarin	Actual Sagarin	Predicted – Actual
2010	Boise State	-0.856	-0.653	1.810	2.463
2010	Air Force	-2.352	-1.795	0.553	2.348
2008	Michigan	1.553	1.185	-0.652	-1.837
2012	Colorado	0.317	0.242	-1.598	-1.840
2008	Washington	0.702	0.536	-1.351	-1.887

Table 3 Quintiles Based on Recruiting Assets

	1	2	3	4	5
Mean					
Recruiting Z	1.380	0.549	–0.012	–0.667	–1.251
Sagarin Z	1.031	0.489	–0.021	–0.606	–0.893
Recruiting r(-1)Z	1.035	0.412	–0.009	–0.500	–0.917
MAE					
Recruiting Z	0.560	0.499	0.544	0.594	0.672
Recruiting r(-1)Z	0.479	0.508	0.544	0.609	0.611
RMSE					
Recruiting Z	0.704	0.627	0.683	0.755	0.873
Recruiting r(-1)Z	0.597	0.632	0.683	0.762	0.775

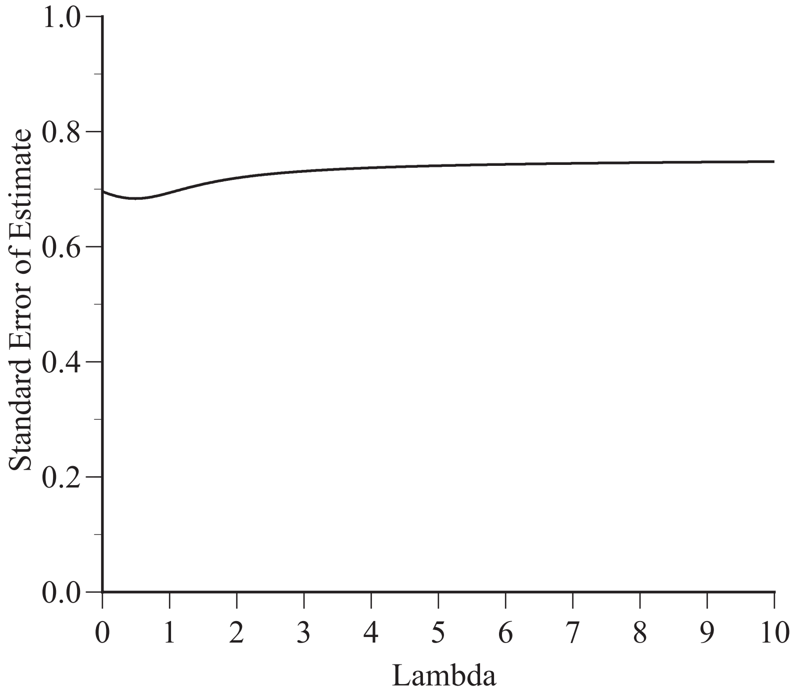

Figure 1 Standard Error of Estimate

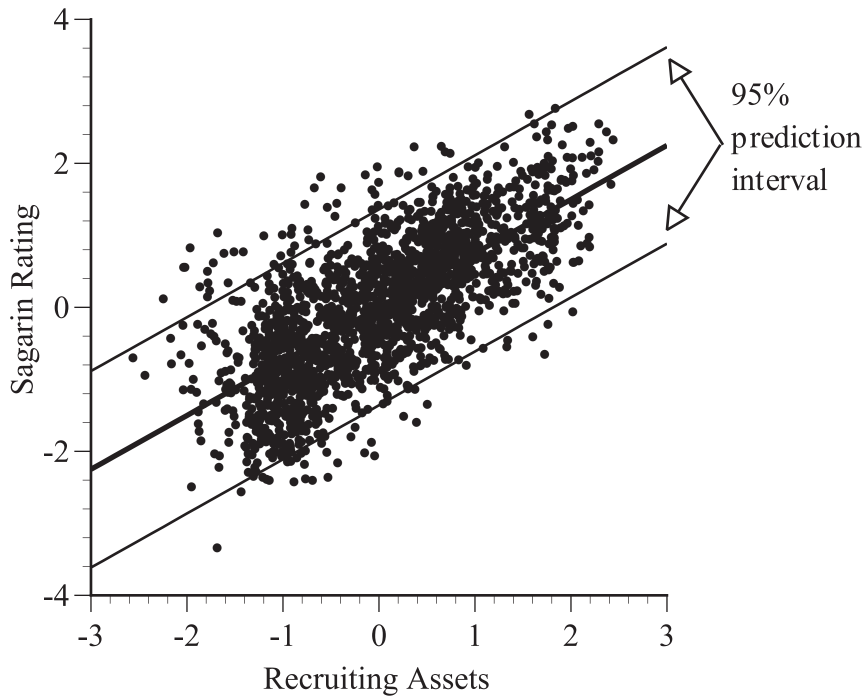

Figure 2 Sagarin Rating and Recruiting Assets