8-2011

NCAA Division I Football Bowl Subdivision: The Importance of Recruiting and Its Relationship with Team Performance

Nathan S. Lloyd
Utah State University

NCAA DIVISION I FOOTBALL BOWL SUBDIVISION:

THE IMPORTANCE OF RECRUITING AND ITS

RELATIONSHIP WITH

TEAM PERFORMANCE

by

Nathan S. Lloyd

A research paper submitted in partial fulfillment

of the requirements for the degree

of

MASTER OF SCIENCE

in

Economics

Approved:

______________________ ______________________

Chris Fawson
Major Professor/Committee Member

Tyler Brough
Committee Member

____

Tyler Bowles
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2011

ABSTRACT

NCAA Division I Football Bowl Subdivision:

The Importance of Recruiting and Its

Relationship with

Team Performance

by

Nathan S. Lloyd, Master of Science

Utah State University, 2011

Major Professor: Dr. Chris Fawson
Department: Economics and Finance

Talent wins college football games. Wins bring in money. Colleges, fans and media hype up the recruiting season as the key to success in the college football season. Is it though? Athletic programs spend large sums of capital and resources to recruit the most talented players possible. This paper explores the relationship between recruited talent and team performance using a simultaneous equations model. Higher players' talent leads to better team performance and a recruiting class has its biggest impact immediately following signing. A team's performance, especially of the most recent season, impacts its ability to recruit. Talent and success experience bidirectional causation, meaning they concurrently cause each other. The theory that top teams maintain top status is true. The theory holds true for all teams as well. Bidirectional causation proved here explains lack of performance mobility across all levels of the Division I Football Bowl Subdivision (FBS).

(38 pages)

ACKNOWLEDGMENTS

I want to thank my wife Danielle for her patience with me as I worked on this research paper and studied for the master of economics degree. She has been a true support. I am grateful for my classmates in this program. We have become good friends and share wonderful memories. Making it through this program was only possible because of their help. I also thank my committee members and major professor. Lastly, I want to thank Utah State University for providing a quality and rigorous education.

Nathan S. Lloyd

CONTENTS

Page

LIST OF TABLES

INTRODUCTION

Despite the amateur status of NCAA athletes, the Division I Football Bowl Subdivision (FBS) (formerly Division IA) is a booming commercial industry. In 2009, the top football programs earned over $50 million in profits (Schwartz, 2009)! Eighty-three of 118 reporting football programs were profitable, with thirty-five of them profiting over $10 million (U.S. Department of Education). Attracting elite high school and junior college players to a program is assumed to translate into future wins, prestige and money for a program. Robert Brown of Cal State University San Marcos estimates that a "premium" player - one later drafted into the NFL-adds $1.1 million of revenue to his university (Brown, 2011).

Every year, college football programs spend huge amounts of time, money and effort in the competitive game off the field known as "recruiting". In 2001, the *average* Division IA program spent $526,000 on recruiting (Weiberg, 2003). This amount jumped to over $750,000 by the 2009 - 2010 academic year, and 29 schools in the FBS spent over $1 million in recruiting alone (U.S. Department of Education). This does not even account for expenses related to recruiting such as scouting and phone bills which are reported to the Department of Education in a separate expense category. It is also important to note that the NCAA limits recruiting activities. The NCAA breaks down the calendar into different periods where varying levels of recruiting can occur. Schools and coaches are limited by quotas in their official visits to prospective student-athletes[1]. Student hosts can be paid a maximum of $30/day to cover entertainment for the host and prospect (NCAA bylaw 13.6.7.5). Advertising for recruits is forbidden (NCAA bylaw 13.4.3.1). These "collusive restrictions on payments" for recruiting

[1] For example, a football program is allowed 56 total official visits annually and the head coach is allowed just one day/year to visit each given prospective student-athlete (see NCAA bylaws 13.6.2.6 and 13.1.2.6.2 for respective examples) (NCAA)

expenses and other factors of football production such as player compensation, are evidence to the NCAA's cartel behavior (Kahn, 2007). If not for these cartel-like behaviors of the NCAA which restrict recruiting, the expense would surely be much greater.

A *Sports Illustrated* article entitled "A History of Recruiting" shows how important football programs perceive recruiting by revealing the great lengths that these programs will go to in order to innovate and work with NCAA bylaws. One school recently produced comic books with their recruit prospects as the main character, leading the team to a national championship. These, along with personalized jerseys, text messaging and other creative recruiting tools have since been banned by the NCAA (Staples, 2008). As of summer 2011, video chats, Facebook accounts and email, among other methods of recruiting, remain legal under NCAA bylaws.

With all the cost and effort surrounding recruiting, one must ask, "Is it worth it? Do better recruits translate into more wins?" Clearly a football team's profit is a function of success on the field but is success a function of its players' talent? When do recruits begin to impact their team and by how much? Does winning increase a team's ability to attract elite recruits?

Answering these critical questions aids athletic administrators and coaches in their jobs. Implications of this study may justify teams in their large expenditures related to recruiting. Results of this study should help teams in their planning and recruiting efforts of players. Knowing they operate in a cycle where recruiting leads to team performance which in turn leads to recruiting, and so forth, would provide insight into needed policy moves. Athletic administrators could focus on breaking the cycle by hiring coaches more focused on recruiting, or more proven as great recruiters, expend more time and money on recruiting and less on stadium/facility enhancement, etc.

LITERARY REVIEW

Sports economics literature involving football falls into various main camps: those modeling production of wins as a function of performance statistics, those modeling attendance as a function of various determinants, literature about coaches, literature involving team performance and its effect on university variables, and those who study the recruiting aspect of the game. An example of the first camp is the work by Keith Willoughby of Bucknell University. He studies winning in the Canadian Football League. The dichotomous dependent variable representing win-loss is regressed on in-game statistics such as the difference in passing yards, rushing yards, turnovers, etc. between the observed team and its opponent. Teams should control the line of scrimmage and focus on dominating these in-game statistics to win the game (Willoughby, 2002). Related studies may differ in their selected explanatory performance statistics used, but most are attempting to model wins in this fashion in order to prescribe coaching policies and emphases for the game. Some studies, such as Stephen Clarke's, look at the effect of home field advantage in athletic competition (Clarke, 2005).

Timothy DeSchriver and Paul Jensen present an economic demand model for spectator attendance in NCAA Division II competitions. Winning percentage and promotional activities such as homecoming positively affect attendance. Winning's effect on attendance grows as the season gets closer to the end (DeSchriver and Jensen, 2002). Attendance has been studied in NCAA Division I as well. A predictive model sets attendance as a function of game-specific, university-specific and team-specific determinants. The recent on-field success of the home team, visiting team, tradition of the home team and being rivals are the biggest positive predictors of attendance demand (Price and Sen, 2003). Conference realignment has an effect on attendance too. Recently, many teams have changed conference affiliation for more competitive conferences. After controlling for the higher quality opponents, their attendance increases (Groza, 2010).

Coaching is another area of football literature. Amy Farmer and Paul Pecorino develop a model to show that under the NCAA's "cartel agreement" to not provide player compensation, coaches' salaries rise (Farmer and Pecorino, 2010). Recently, Paul Holmes presents a model that uses logistic regression methods to estimate dismissal probability for FBS head coaches. Stronger recent team performance decreases the chance of dismissal but stronger historical performances of a team increase the chance of the current coach's dismissal (Holmes, 2011).

One niche in literature specific to college football involves the relationship between team performance and various university variables such as alumni donations, academic quality and state appropriations. "Alumni giving" literature is divided and estimates of the effect a team's success has on donations vary depending on what variables are used in the model, how "success" is defined, and whether the sample includes private or public universities (Kahn, 2007). A football program's culture and tradition, more than on-field success, positively contribute to academic quality (Smith, 2009). Although on-field success does not explain state-government appropriations for universities, simply fielding a football team does (Humphreys, 2006). These are examples of the literature regarding football's effects on campus activity.

Another camp of literature involving college football focuses on the aspect of recruiting. Klenosky, Templin and Troutman of Purdue University present an economic model of the college football recruiting process (Klenosky, et. al, 2001). They find that factors such as the coaching staff, playing time potential and playing on television have some significance in determining the recruit's school of choice. Dumond, Lynch and Platania model recruits as rational agents seeking to maximize their discounted expected utility. They create a predictive model to determine the likelihood a recruit will sign with a given school. Factors influencing a recruit's decisions, conditional on being offered the scholarship, include the distance between the school's city and his hometown, whether or not the team is a "BCS" team, the team's recent performance, academic reputation and media exposure (Dumond, et. al, 2008).

This project aligns with a different aspect of recruiting literature however- exploring the relationship between recruiting and team performance. George Langelett finds that teams with success on the field are able to attract quality recruits, which in turn increases the quality of future team performance. Only the top 10 recruiting classes are observed along with the top 25 teams for each year in his study; the data set he uses is not representative of Division I FBS (Langelett, 2003). His feedback system between team performance and recruiting supports the phenomenon where top teams remain top teams. My study seeks to test the generalized conclusions of Langelett. I test if the immobile nature (in terms of performance) of the vast majority of teams across all FBS levels can be explained by the theory of bidirectional causation of recruiting and team performance.

To show the lack of mobility of teams across performance levels, consider the following facts: the net average change for a team's rating from year to year is .021; for the median team of 2010, taking the average change would translate into the *exact same* ranking in 2011 (given the same team ratings of 2010 for the other teams). The largest rating gain from year-to-year was 28.73 in my data set- even the bottom 48.7% of 2010 teams could not be ranked number 1 next year with that kind of enormous improvement. Also, in 13 years only 43 unique FBS teams have ever finished in the top 10[2]. Lack of performance mobility may be explained by bidirectional causation of team performance and recruiting.

---

[2] Author's data calculations

THEORETICAL FRAMEWORK

The buzz and hoopla surrounding National Signing Day[3] and the entire recruitment season is based on the assumption that a talented recruiting class will convert into a high-ranked football team over the next four years. George Langelett showed this assumption to be true in his 2003 article, published in the *Journal of Sports Economics*. I hypothesize this theory should be extended to the entire FBS. If top talent leads to high team performance, the opposite should be true - low talent should produce lower ranked teams. Theory suggests recruits affect team performance most their freshman year (year signed=red-shirt year, next year = freshman year), but the impact of a recruit is discounted over the remainder of his time with the program (Langelett, 2003).

Another common perception is that teams with recent success will attract better recruits. Specifically, theory states that a team's performance best explains recruiting at its two-year lag (Langelett, 2003). This bidirectional theoretical framework of team performance and recruiting will be modeled empirically.

[3] First Wednesday in February

ECONOMETRIC MODELING

I propose a simultaneous equations model to test the theoretical hypotheses, because of the bidirectional nature of team performance and recruiting. I take advantage of recruiting class data and team rankings now available for all teams in the FBS. Following literature, the first equation (Equation I) includes Team Performance (TP) which is regressed solely on Recruiting Class talent (RC) and its lags. Since players are given five years of eligibility when using a red-shirt year, they may affect a team's performance over five years. The fifth-year lag of recruiting classes is included in the model to control for this possibility. Equation I appears as:

Team Performance = f (Red-shirt, Freshman, Sophomore, Junior, Senior),

or, alternatively in lag form:

Team Performance = f (RC, $\mathbf{RC_{t-1}}$, $\mathbf{RC_{t-2}}$, $\mathbf{RC_{t-3}}$, $\mathbf{RC_{t-4}}$).

Team Performance is the dependent variable of Equation I. Using Jeff Sagarin's computer ratings of all FBS teams, teams' performance is measured by their end-of-season rating. Like Groza, I use Sagarin ratings to represent recent on-field success (Groza, 2010). Sagarin ratings are used to measure team performance because of their validity in the sports world, being the supplier to the USA Today's rankings. Also, the Bowl Championship Series (BCS) uses Sagarin ratings to formulate their rankings and make decisions on their prestigious bowl game participants. Other popular polls such as the AP Top 25 or BCS rankings are not used since these only rank the top teams, and not every team in the FBS. Team ratings are used, as opposed to rankings, for accuracy of measurement since the difference between pairs of rankings is not equidistant, and to avoid censored data.

Recruited talent of a team is represented by the recruiting ranking each entering class receives upon joining a team. Every year, scout.com rates all players who sign with NCAA FBS teams on a scale of 1-5 “stars”. Recruits of the highest talent are given 5 stars, followed by those with 4 stars, etc. down to 1 for the least-talented recruits. Scout.com awards points to every team, taking into account both their recruits’ absolute talent and their relative talent to other teams[4]. Rankings are assigned each FBS team as determined by their overall point total[5]. Scout.com was acquired by Fox Interactive Company in 2005 to be a supplier of information to Fox Sports. Scout.com and its affiliate in production, Superprep.com, are leaders of the recruiting information industry.

Equation I is represented algebraically as follows:

$$\boldsymbol{Y_{i,t} = \alpha_i + \beta_1 X_{i,t} + \beta_2 X_{i,t-1} + \beta_3 X_{i,t-2} + \beta_4 X_{i,t-3} + \beta_5 X_{i,t-4} + \varepsilon_{it}}$$

where Y is the final team rating, i indexes the team of analysis, t is the index for time, α is the intercept term, ε is the error term to account for the random nature of models, and β is the coefficient for explanatory variable X, the team’s recruiting class ranking (RC). RC and its four lags are included in the model to cover the time athletes play for a team. β_1 is the beta coefficient for the recruiting class of the current year, or the red-shirt freshmen. β_2 is the beta coefficient for the recruiting class of last year, or the current freshman players, and so forth.

Because recruiting is also affected by prior performance of a team, the empirical model is only accurate with the inclusion of Equation II. This tests the theory that a high school player uses current teams’ performance to choose where to attend school. Dumond, Lynch and Platania showed that a team’s current performance is a major factor in a recruit’s decision of where to play in college (Dumond, et. al, 2008). Equation II is given as:

---

[4] The formula is based on a player’s rating and ranking: 5-star=200 points, 4-star=120 points, 3-star=40 points, 2-star=20 points, 1-star=0 points. The number 1-ranked player (assuming 100 players in his position) of a position=100 points, number 2=99 points, down to number 100=1 point. A maximum of 25 recruits/team are evaluated towards a team’s points and ranking.

[5] These rankings are updated regularly and slightly change from time to time. My data was last updated July 2011.

Recruiting Class = g (Senior HS, Junior HS, Sophomore HS, Freshman HS),

which allows for the player's four years of high school attendance and his decision period. This is shown alternatively with team performance lags:

Recruiting Class = g ($\boldsymbol{TP_{t-1}, TP_{t-2}, TP_{t-3}, TP_{t-4}}$)

Algebraically, this becomes:

$$\boldsymbol{Y_{i,t} = \alpha_i + \beta_1 X_{i,t-1} + \beta_2 X_{i,t-2} + \beta_3 X_{i,t-3} + \beta_4 X_{i,t-4} + \varepsilon_{it}}$$

where Y is the recruiting class ranking, i indexes the team of analysis, t is the index for time, α is the intercept term, ε is the error term to account for the random nature of models, and β is the coefficient for explanatory variable X, the team's performance (TP). Four lags are included in the model to cover the time recruits attend high school and scout their potential future teams. β_1 is the beta coefficient for the team performance last year, or while the recruits are seniors in high school. β_2 is the beta coefficient for the team's performance two years ago, or while the recruits are juniors in high school, and so forth.

All teams in the FBS are observed for the data set. The cross section of teams is studied over time, using data from 1998 to the present. The frequency of data is yearly- for each season. With 120 teams under observation for 14 years, 1649 total observations will be included in this data set with the final year of data being incomplete[6]. Explanatory variables of Equation I are observed for 2011, but not the dependent variable. Recruiting rankings are observable from 2002 on, while team ratings are available since 1998. With four lags needed in Equation I, there is enough data to begin the data series in 2006, lasting six years, ending with 2011 inclusive. For Equation II, there is enough data to begin the data series in 2002. This panel

[6] 8 schools unbalance the panel since their teams are newer than 1998. Buffalo and Middle Tennessee began in 1999, UConn in 2000, South Florida and Troy in 2001, FAU and Florida International in 2004, and Western Kentucky in 2007. They are not thrown out of the data in order to keep the results unbiased (Most these teams are from the bottom half of the team ratings).

study gleans data from various sources: Scout.com, Sports-reference.com, ESPN.go.com and USA Today[7].

To summarize, the empirical model (Model I) of simultaneous equations is given as:

Equation I: TP = f (RC, $\mathbf{RC_{t-1}}$, $\mathbf{RC_{t-2}}$, $\mathbf{RC_{t-3}}$, $\mathbf{RC_{t-4}}$)

Equation II: RC = g ($\mathbf{TP_{t-1}}$, $\mathbf{TP_{t-2}}$, $\mathbf{TP_{t-3}}$, $\mathbf{TP_{t-4}}$)

All the variables used are displayed in the table below:

List of Variables	
Variable:	**Represents:**
TP (Dependent)	Team performance- Final Sagarin rating of the year
RC (Explanatory in Equation I, Dependent in Equation II)	Recruiting Class Ranking (#1 is best) of current red-shirt players
RC, t-1 (Explanatory in Equation I, along with remaining RC lags)	RC's 1st lag or current freshman players
RC, t-2	RC's 2nd lag or current sophomore players
RC, t-3	RC's 3rd lag or junior players
RC, t-4	RC's final lag or senior players
TP, t-1 (Explanatory in Equation II, as are the remaining variables)	Team Performance's 1st lag
TP, t-2	Team Performance's 2nd lag
TP, t-3	Team Performance's 3rd lag
TP, t-4	Team Performance's 4th lag

[7] http://www.usatoday.com/sports/sagarin-archive.htm

Team Performance and Recruiting Class Rankings have the following descriptive statistics:

Summary Statistics[8]			
TP		**RC**	
Mean	70.8236	**Mean**	59.1121
Standard Error	0.3144	**Standard Error**	0.9860
Median	71.23	**Median**	59
Mode	65.9	**Mode**	107
Standard Deviation	12.2956	**Standard Deviation**	33.7130
Sample Variance	151.1806	**Sample Variance**	1136.5670
Kurtosis	-0.30	**Kurtosis**	-1.1766
Skewness	-0.0930	**Skewness**	0.0038
Range	76.46	**Range**	119
Minimum	30.47	**Minimum**	1
Maximum	106.93	**Maximum**	120
Count	1529	**Count**	1190

[8] All output values throughout the paper are rounded to the nearest ten-thousandth

ESTIMATION

Estimation of the empirical model is performed with the instrumental variable technique using the Three-Stage Least Squares (3SLS) and Two-Stage Least Squares (2SLS) regression methods. This is accomplished using *R* software and its “Systemfit” package, which includes functions for simultaneous equations and testing data. The 3SLS and 2SLS estimators are chosen as opposed to Ordinary Least Squares (OLS) because the OLS estimator is biased in simultaneous equations (Hamann and Henningsen, 2007). The disturbance term of one equation and a regressor are correlated[9], violating Assumption Three (exogeneity of the independent variables) of the Classical Linear Regressions Model (Greene, 2008). Weighted Least Squares (WLS) and Seemingly Unrelated Regression (SUR) are not estimators of choice either since they rely on the assumption of exogeneity.

Given that variable RC is an endogenous explanatory variable in Equation I, the instrumental variable technique is used for estimation. A proper choice for an instrumental variable is one that is not correlated to the disturbance term, but is correlated with the endogenous regressor (Greene, 2008). To instrument for endogenous variables in simultaneous equations, exogenous variables are used (Hamann and Henningsen, 2007). For this model, RC and its four lags along with TP’s four lags are all proper instrument choices. Langelett used only 3SLS estimation, but 3SLS here tests negative for consistency using the Hausman Specification Test, shown in the next table (Langelett, 2003). Under the null hypothesis of this test, all exogenous variables are uncorrelated with all disturbance terms (Hamann and Henningsen, 2007). Therefore, 3SLS is inconsistent and 2SLS is the preferred estimator for this study.

[9] Precisely, the Equation II disturbance term is correlated with Equation II’s RC since RC is an endogenous variable. When the disturbance is high, RC is high too, which simultaneously raises RC in equation I.

Hausman Specification Test	
Data	Model I
Hausman Test Statistic	171.1664
Degrees of Freedom	11
P-value	<2.2e-16

Model I's output[10] is summarized below:

Model I Output (2SLS)[11]				
Equation I Adj. R-squared: .4846	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
Intercept	85.7842	0.7190	119.3181	***
RC	-0.0833	0.0239	-3.4846	***
RC_1	-0.0475	0.0247	-1.9258	.
RC_2	-0.0787	0.0257	-3.0596	**
RC_3	-0.0515	0.0253	-2.0375	*
RC_4	0.0118	0.0239	0.4931	
Significance Code	*** .001	** .01	*.05	. .10
Equation II Adj. R-squared: .5984	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
Intercept	240.9283	6.1835	38.9632	***
TP_1	-0.8621	0.1231	-7.0013	***
TP_2	-0.4894	0.1329	-3.6830	***
TP_3	-0.5306	0.1318	-4.0247	***
TP_4	-0.6778	0.1193	-5.6798	***
Significance Code	*** .001	** .01	*.05	. .10

[10] The estimation of systems of equations with unequal numbers of observations has not been thoroughly tested yet, therefore lowering the n observations to 591 (Hamann and Henningsen, 2007)

[11] See Appendix A for output results using the 3SLS estimator

All variables are statistically significant and their coefficient estimates have the expected negative sign, with the exception of RC_4. This exception is an inaccurate result that is fixed in Model II. For Equation I and II, recruiting class rankings are set up so that lower is better. The top recruiting class each year is awarded the #1 ranking. The worst recruiting class is ranked #120[12]. Higher ratings for teams indicate better performance. Therefore, for better recruiting classes to lead to higher team performance, the negative sign is expected. For recent on-field success to attract more talented recruits, the negative coefficient sign is also correct.

The model tests negative for heteroscedasticity using the Studentized Breusch-Pagan Test. I fail to reject the null hypothesis of homoscedasticity as seen in the following table:

Studentized Breusch-Pagan Test		
Data	Equation I	Equation II
BP test statistic	2.32	2.3081
Degrees of Freedom	5	4
P-value	.8031	.6793

However, the model tests positive for multicollinearity. The following tables show the correlation between variables which indicates the problem of collinearity:

[12] 120 teams have existed since 2007, 119 from 2004-2006, 117 from 2001-2003, 115 in 2000, 114 in 1999, and 112 in 1998

Equation I Correlation[13]	TP	RC	RC_1	RC_2	RC_3	RC_4
TP	1	-.67	-.66	-.67	-.66	-.62
RC	-.67	1	.87	.87	.86	.85
RC_1	-.66	.87	1	.87	.86	.85
RC_2	-.67	.87	.87	1	.88	.86
RC_3	-.66	.86	.86	.88	1	.87
RC_4	-.62	.85	.85	.86	.87	1

Equation II Correlation	RC	TP_1	TP_2	TP_3	TP_4
RC	1	-.68	-.68	-.67	-.66
TP_1	-.68	1	.75	.66	.61
TP_2	-.68	.75	1	.74	.66
TP_3	-.67	.66	.74	1	.73
TP_4	-.66	.61	.66	.73	1

An absolute value of 0.8 or greater correlation between two variables is considered high collinearity, while .5 - .8 is the range used for moderate collinearity (Kennedy, 2008). The high multicollinearity in Equation I explains the strange exception found in this equation. To correct for multicollinearity, OLS regressions are run on each explanatory variable, following literary practice (Langelett, 2003). These results are shown at continuation and summarized as "Model II":

[13] Figures in correlation tables rounded to nearest one-hundredth

Model II Output (Separate Regressions)						
Equation I: Dependent Variable – TP	**Adj. R-squared**	**Constant**[14]	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
RC	.4423	84.6850	-0.2305	.0080	-28.85	***
RC_1	.4270	84.4033	-0.2259	.0085	-26.43	***
RC_2	.4276	84.3528	-0.2244	.0090	-24.80	***
RC_3	.4354	84.5547	-0.2264	.0010	-23.37	***
RC_4	.3887	83.7198	-0.2133	.0110	-19.40	***

Equation II: Dependent Variable – RC	**Adj. R-squared**	**Constant**	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
TP_1	.4688	197.0505	-1.9434	0.0605	-32.12	***
TP_2	.4636	193.9419	-1.9021	0.0598	-31.79	***
TP_3	.4433	189.1465	-1.8346	0.0601	-30.51	***
TP_4	.4394	185.2543	-1.7810	0.0588	-30.27	***
Significance	*** .001	** .01	*.05	. .10		

RC_4 in the Team Performance Equation now has the expected negative sign for its coefficient estimate and all explanatory variables in both equations are highly significant. The most recent class to sign has the strongest impact on team performance. This will include freshmen, but mostly red-shirt players. This contradicts the hypothesis suggesting that the first lag, or mostly freshman players, has the strongest effect. The recruiting class with the second highest impact (by evaluating coefficient estimates) on team performance is the third lag, or mostly junior players. This result is intuitive because recruits should grow and improve in their coach's system, and by their final years be playing more than the other players. With junior

[14] See Appendix B for the constants' standard error, T-Value and significance level values

players (and senior players who have not red-shirted) producing more minutes played, their talent should have a greater impact on their team's performance than the previous two lags.

Team Performance is found to significantly explain recruiting. Contrary to the hypothesis, TP_1, or a team's performance while a recruit is a senior in high school, has the strongest impact on recruiting. This result is instinctive, since high school recruits make final decisions on where to play during their senior year. High school recruits appear to base their decision off of the most recent on-field results.

CONCLUSIONS AND SHORTCOMINGS

These models show a “feedback” system in the college football market. More work could be done to study the Team Performance Equation. Theory states that the freshmen (recruiting class talent lagged one year) most impact a team’s performance, while this study found the first year recruiting class to most impact the team’s success. Perhaps less athletes are red-shirting than when George Langelett studied this issue. Maybe the composition of recruits is changing with the junior college & transfer/high school recruit ratio rising. I believe talent’s role in the college football market, including coaching talent, needs to be better understood. Because recruiting classes experience their strongest effect on team performance right after signing, further study needs to determine if coaching talent is explaining team performance after the first year. Other factors besides coaching talent may explain the diminished recruiting effect on team performance. Perhaps the model should be expanded to include other variables.

Schools in the FBS expend significant amounts of resources on recruiting. Equation I proves that recruiting impacts team performance. Schools are validated in their actions by this evidence. Recruits significantly impact their team years after signing, with their impact being largest the first year. Team performance in turn affects recruiting. Prospective FBS football players appear to base their decision of where to play mainly on the most recent college football results, while they are seniors in high school. The bidirectional causation theory of recruiting talent and team performance extends from its application of top teams to the entire FBS and explains the lack of team performance mobility from year to year.

REFERENCES

Brown, R. (2011). Research Notes: Estimates of College Football Player Rents. *Journal of Sports Economics,* 12(2), 200-212. Retrieved April 2011 from Sage Publications Database. Online version found at http://jse.sagepub.com/content/12/2/200.

Clarke, S.R. (2005). Home Advantage in the Australian Football League. *Journal of Sports Sciences,* 23(4), 375-385. Retrieved April 2011 from Academic Search Premier Database. Online version found at http://search.ebscohost.com.dist.lib.usu.edu/login.aspx?direct=true&db=aph&AN=16968656&site=ehost-live.

DeSchriver, T.D.; Jensen, P.E. (2002). Determinants of Spectator Attendance at NCAA Division II Football Contests. *Journal of Sport Management*, 16, 311-330. Retrieved May 2011 from Econ Lit Database.

Dumond, J.M.; Lynch, A.K.; Platania, J. (2008). An Economic Model of the College Football Recruiting Process. *Journal of Sports Economics*, 9(1), 67-87. Retrieved May 2011 from Econ Lit Database.

Farmer, A. and Pecorino, P. (2010). Is the Coach Paid Too Much?: Coaching Salaries and the NCAA Cartel. *Journal of Economics and Management Strategy*, 19(3), 841-862. Retrieved June 2011 from Econ Lit Database.

Greene, W.H. (2008). Econometric Analysis. *Pearson Prentice Hall*. Edition 6.

Groza, M.D. (2010). NCAA Conference Realignment and Football Game Day Attendance. *Managerial and Decision Economics*, 31(8), 517-529. Retrieved June 2011 from Econ Lit Database.

Hamann, J.D.; Henningsen, A. (2007). Systemfit: A Package for Estimating Systems of Simultaneous Equations in R. *Journal of Statistics Software*, 2007. Retrieved January 2011 from *Rstudio*.

Holmes, P. (2011). Win or Go Home: Why College Football Coaches Get Fired. *Journal of Sports Economics,* 12(2), 157-178. Retrieved June 2011 from Econ Lit Database.

Humphreys, B.R. (2006). The Relationship between Big-Time College Football and State Appropriations for Higher Education*. International Journal of Sport Finance*, 1(2), 119. Retrieved June 2011 from Econ Lit Database.

Kahn, L. M. (2007). Markets: Cartel Behavior and Amateurism in College Sports. *Journal of Economic Perspectives*, 21(1), 209-226. Retrieved June 2011 from Econ Lit Database.

Kennedy, P. (2008). A Guide to Econometrics. *Blackwell Publishing.* Edition 6.

Klenosky, D.B.; Templin, T.J.; Troutman, J.A (2001). Recruiting Student Athletes: A Means-End Investigation of School-Choice Decision Making. *Journal of Sport Management,* 15(2), 95-106. Retrieved March 2011 from Academic Search Premier Database.

Langelett, G. (2003). The Relationship between Recruiting and Team Performance in Division 1A College Football. *Journal of Sports Economics*, 4(3), 240-245. Retrieved March 2011 from Sage Publications Database. Online version found at http://jse.sagepub.com/content/4/3/240.

NCAA. (2010). 2010-11 NCAA Division I Manual. The *National Collegiate Athletic Association.* Retrieved July 2011 from http://www.ncaapublications.com/productdownloads/D111.pdf.

Price, D. I.; Sen, K.C. (2003). The Demand for Game Day Attendance in College Football: An Analysis of the 1997 Division 1-A Season. *Managerial and Decision Economics*, 24(1), 35-46. Retrieved June 2011 from JSTOR Database.

Schwartz, P.J. (2009). College Football's Most Valuable Teams. *Forbes.com.* Retrieved April 2011 from http://www.forbes.com/2009/12/22/most-vaulable-college-football-teams-business-sports-college-football.html.

Smith, D. R. (2009). College Football and Student Quality: An Advertising Effect or Culture and Tradition? *American Journal of Economics & Sociology*, 68(2), 553-579. Retrieved July 2011 from Econ Lit Database.

Staples, A. (2008). A History of Recruiting; How Coaches Have Stayed a Step Ahead. *Sports Illustrated*, June 23. Retrieved July 2011 from http://sportsillustrated.cnn.com/2008/writers/andy_staples/06/19/recruiting.main/index.

U.S. Department of Education. (n.d.). The Equity in Athletics Data Analysis Cutting Tool. *Office of Postsecondary Education*. Retrieved June 2011 from http://ope.ed.gov/athletics/.

Wieberg, S. (2003). Big spending no guarantee of big revenue. *USA Today,* August 14. Retrieved June 2011 from http://www.usatoday.com/sports/college/2003-08-14-spending-revenue-study_x.htm.

Willoughby, K. (2002). Winning Games in Canadian Football: A Logistic Regression Analysis. *College Mathematics Journal,* 33(3), 215-220. Retrieved March 2011 from http://www.edwards.usask.ca/faculty/Keith%20Willoughby/files/cmj215-220.pdf.

Appendix A – Model I Output (3SLS)

Equation I Adj. R-squared: .4606	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
Intercept	87.9945	.7085	124.2045	***
RC	-0.1816	.0227	-8.0137	***
RC_1	-0.0295	.0234	-1.2617	
RC_2	-0.0624	.0244	-2.5600	*
RC_3	-0.0340	.0240	-1.4161	
RC_4	0.0207	.0226	.9153	
Significance Code	*** .001	** .01	*.05	. .10
Equation II Adj. R-squared: .5952	**Coefficient Estimates**	**Standard Error**	**T-Value**	**Significance**
Intercept	247.6817	6.0701	40.8034	***
TP_1	-1.1067	.1169	-9.4668	***
TP_2	-.4454	.1260	-3.5333	***
TP_3	-.5090	.1251	-4.0698	***
TP_4	-0.5934	.1133	-5.2368	***
Significance Code	*** .001	** .01	*.05	. .10

Appendix B – Model II's Output of Intercepts

Equation I: Dependent Variable – TP	**Adj. R-squared**	**Constant**	**SE**	**T-Value**	**Significance**
RC	.4423	84.6850	.5423	156.17	***
RC_1	.4270	84.4033	.5810	145.28	***
RC_2	.4276	84.3528	.6160	136.9	***
RC_3	.4354	84.5547	.6620	127.72	***
RC_4	.3887	83.7198	.7535	111.10	***
Equation II: Dependent Variable – RC	**Adj. R-squared**	**Constant**	**SE**	**T-Value**	**Significance**
TP_1	.4688	197.0505	4.3544	45.25	***
TP_2	.4636	193.9419	4.3023	45.08	***
TP_3	.4433	189.1465	4.3247	43.74	***
TP_4	.4394	185.2543	4.2318	43.78	***

Appendix C – R Code

```
## Counting profits, recruiting expenses...

setwd("C:\\Users\\Nate\\Desktop")

dat <- read.csv("thesis.csv", header=T)

head(dat)

n<-nrow(dat)

dat$profit <- rep("", n)

head(dat)

for (i in 1:n) {

  if (dat$Revenues[i] > dat$Expenses[i])

  dat$profit[i]<-1

  else dat$profit[i]<-0

}

head(dat)

table(dat$profit)

dat$prof.amount<-dat$Revenues - dat$Expenses

head(dat)

dat

dat$profit.10m <- rep("",n)

head(dat)

for (i in 1:n) {

  if (dat$prof.amount[i] > 9999999)

  dat$profit.10m[i]<-1

  else dat$profit.10m[i]<-0

}
```

```
head(dat,20)
table(dat$profit.10m)
max(dat$prof.amount)
outfile <- "C:\\Users\\Nate\\Desktop\\footballprofit.csv"
write.table(dat, file=outfile, quote=FALSE, sep=",", row.names=FALSE, col.names=TRUE)


## For recruiting expenses
dat <- read.csv("ADprofit.csv", header=T)
head(dat)
n<-nrow(dat)
dat$RE.1m <- rep("", n)
head(dat)
for (i in 1:n) {
  if (dat$Total[i] > 999999)
  dat$RE.1m[i]<-1
  else dat$RE.1m[i]<-0
}
head(dat)
table(dat$RE.1m)
## Save together
dat$prof.amount<-dat$Revenues - dat$Expenses
head(dat)
dat$profit.10m <- rep("",n)
head(dat)
for (i in 1:n) {
```

```
 if (dat$prof.amount[i] > 9999999)
 dat$profit.10m[i]<-1
 else dat$profit.10m[i]<-0
}
head(dat,20)
table(dat$profit.10m)
max(dat$prof.amount)
outfile <- "C:\\Users\\Nate\\Desktop\\ADprofit.csv"
write.table(dat, file=outfile, quote=FALSE, sep=",", row.names=FALSE, col.names=TRUE)
## Repeated for csv with 2003 figures

## Code for getting variable absolute difference:
setwd("C:\\Users\\Nate\\Desktop")
dat<-read.csv("Sagarin.csv", header=T)
head(dat)
dat<-dat[1:9]
head(dat)
dat$Difference <- abs(dat$Difference)
head(dat)
outfile <- "C:\\Users\\Nate\\Desktop\\Sagarin.csv"
write.table(dat, file=outfile, quote=FALSE, sep=",", row.names=FALSE, col.names=TRUE)

## Code for obtaining range variable:

rang<-list()
```

```
for (i in 1:120) {
  rang[[i]] <- range(tmpdat[[i]]$Rating)
}
rang


## Equation I code of Simultaneous System
# Creating lags for RC ranking
setwd("C:\\Users\\Nate\\Desktop")
dat <- read.csv("sagarin.csv", header=T)
head(dat)
dat<-dat[1:11]
head(dat)
teams <- as.character(unique(dat$Team))
teams
length(teams)
tmp <- list()
dat$Team <- as.character(dat$Team)
dat$Conf.11<- as.character(dat$Conf.11)
tmpdat <- list()
for(i in 1:120) { tmpdat[[i]] <- dat[dat$Team == teams[i], ]; }
tmpdat[[1]]
newdat <- list()
tmp <- tmpdat[[1]]
lags <- 4
```

```
for(i in 1:120) { tmp <- tmpdat[[i]]; n <- nrow(tmp); newdat[[i]] <- cbind(tmp$Rating[(lags+1):n],
        tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n],embed(tmp$RC, lags+1)); }
newdat
newdat <- do.call("rbind", newdat)
newdat <- list()
for(i in 1:120) { tmp <- tmpdat[[i]]; n <- nrow(tmp); newdat[[i]] <- cbind(tmp$Rating[(lags+1):n],
        tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n], embed(tmp$RC, lags+1)); }
names(newdat) <- teams
for(i in 1:120) { n <- nrow(newdat[[i]]); newdat[[i]] <- cbind(rep(names(newdat[[i]]), n), newdat[[i]]) }
newdat
data.class(names(newdat))
tmp
n <- nrow(tmp)
new <- cbind(tmp$Rating[(lags+1):n], tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n],embed(tmp$RC,
        lags+1))
nm <- unique(as.character(tmp$Team))
n <- nrow(new)
nate <- data.frame(cbind(rep(nm, n), new))
newdat <- list()
names(tmpdat) <- teams
tmpdat

for(i in 1:120) { n <- nrow(tmpdat[[i]]); newdat[[i]] <- data.frame(Team=rep(teams[i], n-lags),
        Rating=tmpdat[[i]]$Rating[(lags+1):n], Year=tmpdat[[i]]$Year[(lags+1):n],
        Conf.11=tmpdat[[i]]$Conf.11[(lags+1):n],embed(tmpdat[[i]]$RC, lags+1)); }
```

```
newdat <- do.call("rbind", newdat)

newdat

names(newdat) <- c("Team", "Rating", "Year", "Conf.11", "RC", "RC_1", "RC_2", "RC_3", "RC_4")

head(newdat,25)

outfile <- "C:\\Users\\Nate\\Desktop\\Model4.csv"

write.table(newdat, file=outfile, quote=FALSE, sep=",", row.names=FALSE, col.names=TRUE)


#####  Equation II of Simultaneous Eq system

# Creating lags for Team Performance (Rating)

setwd("C:\\Users\\Nate\\Desktop")

dat <- read.csv("sagarin.csv", header=T)

dat<-dat[1:11]

head(dat)

teams <- as.character(unique(dat$Team))

teams

length(teams)

tmp <- list()

dat$Team <- as.character(dat$Team)

dat$Conf.11<- as.character(dat$Conf.11)

tmpdat <- list()

for(i in 1:120) { tmpdat[[i]] <- dat[dat$Team == teams[i], ]; }

tmpdat[[1]]

newdat <- list()

tmp <- tmpdat[[1]]

lags <- 4
```

```
for(i in 1:120) { tmp <- tmpdat[[i]]; n <- nrow(tmp); newdat[[i]] <- cbind(tmp$RC[(lags+1):n],
        tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n],embed(tmp$Rating, lags+1)); }
newdat
newdat <- do.call("rbind", newdat)
newdat <- list()
for(i in 1:120) { tmp <- tmpdat[[i]]; n <- nrow(tmp); newdat[[i]] <- cbind(tmp$RC[(lags+1):n],
        tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n], embed(tmp$Rating, lags+1)); }
names(newdat) <- teams
for(i in 1:120) { n <- nrow(newdat[[i]]); newdat[[i]] <- cbind(rep(names(newdat[[i]]), n), newdat[[i]]) }
newdat
data.class(names(newdat))
tmp
n <- nrow(tmp)
new <- cbind(tmp$RC[(lags+1):n], tmp$Year[(lags+1):n], tmp$Conf.11[(lags+1):n],embed(tmp$Rating,
        lags+1))
nm <- unique(as.character(tmp$Team))
n <- nrow(new)
nate <- data.frame(cbind(rep(nm, n), new))
newdat <- list()
names(tmpdat) <- teams
tmpdat

for(i in 1:120) { n <- nrow(tmpdat[[i]]); newdat[[i]] <- data.frame(Team=rep(teams[i], n-lags),
        RC=tmpdat[[i]]$RC[(lags+1):n], Year=tmpdat[[i]]$Year[(lags+1):n],
        Conf.11=tmpdat[[i]]$Conf.11[(lags+1):n],embed(tmpdat[[i]]$Rating, lags+1)); }
```

```
newdat <- do.call("rbind", newdat)

newdat

names(newdat) <- c("Team", "RC", "Year", "Conf.11", "Rating", "Rating_1", "Rating_2", "Rating_3",
        "Rating_4")

head(newdat,25)

outfile <- "C:\\Users\\Nate\\Desktop\\Model5.csv"

write.table(newdat, file=outfile, quote=FALSE, sep=",", row.names=FALSE, col.names=TRUE)


## Model II contains everything to now run Simultaneous EQ Code: 3SLS

setwd("C:\\Users\\Nate\\Desktop")

dat <- read.csv("Model2.csv", header=TRUE)

library(systemfit)

data.frame(dat)

dat1 <- na.omit(dat)

library(plm)

dat2 <- pdata.frame(dat1, index=c("Team"), drop.index=TRUE, row.names=TRUE)

head(dat2)


I <- Rating ~ RC + RC_1 + RC_2 + RC_3 + RC_4

II <- RC ~ Rating_1 + Rating_2 + Rating_3 + Rating_4

inst <- ~RC + RC_1 + RC_2 + RC_3 + RC_4 + Rating_1 + Rating_2 + Rating_3 + Rating_4

system <- list(I=I, II=II)

fit3sls <- systemfit(system, method = "3SLS", inst = inst, data = dat2)

summary(fit3sls)

fit2sls <- systemfit(system, method = "2SLS", inst = inst, data=dat2)
```

```
summary(fit2sls)

## NB: Pooled by default = false. Same results as if pooled=TRUE in system fit function


## Hausman Specification test:

h<-hausman.systemfit(fit2sls, fit3sls)

print(h)

## Reject the Null hypothesis - 3SLS is inconsistent


## Test for Heteroskedasticity

bptest(I, data=dat)

bptest(II, data=dat)

## Homoskedasticity assumption holds


## Test for multicollinearity: EQ I

T<-(cbind(dat$Rating, dat$RC, dat$RC_1, dat$RC_2, dat$RC_3, dat$RC_4))

T<-as.matrix(T)

T<-na.omit(T)

cor(T)

## Very Large collinearity (>.8)

## Multicollinearity: EQ II

T<-(cbind(dat$RC, dat$Rating_1, dat$Rating_2, dat$Rating_3, dat$Rating_4))

T<-as.matrix(T)

T<-na.omit(T)

cor(T)

## Moderate Collinearity (<.8)
```

```
## Model III: separate regression for each variable:

dat <- read.csv("Model2.csv", header=TRUE)

mod1.RC <- lm(Rating ~ RC, data=dat)

summary(mod1.RC)

mod1.RC_1 <- lm(Rating ~ RC_1, data=dat)

summary(mod1.RC_1)

mod1.RC_2 <- lm(Rating ~ RC_2, data=dat)

summary(mod1.RC_2)

mod1.RC_3 <- lm(Rating ~ RC_3, data=dat)

summary(mod1.RC_3)

mod1.RC_4 <- lm(Rating ~ RC_4, data=dat)

summary(mod1.RC_4)


## Equation II:

Rating1 <- lm(RC ~ Rating_1, data=dat)

summary(Rating1)

Rating2 <- lm(RC ~ Rating_2, data=dat)

summary(Rating2)

Rating3 <- lm(RC ~ Rating_3, data=dat)

summary(Rating3)

Rating4 <- lm(RC ~ Rating_4, data=dat)

summary(Rating4)
```