Senior Project
Department of Economics

"Do the Fans in the Seats React to the Teams on the Field?"

William Gilkison
May 2011

Advisors: *Francesco Renna*

Table of Contents

Abstract

NCAA FBS football teams hosted over 37 million fans in the 2008-2009 season, 20 million more fans than the NFL. Division IA schools averaged over 45,000 fans per game, but seasonal measures of competitive balance cannot explain the change in attendance at the game time level. The quality of teams playing at game time results in a change in attendance. The competitive balance between the teams is significant in understanding the change in demand for tickets, as well as the probability a stadium will be sold out. This is the first paper providing discussion on game time attendance for NCAA football and first paper to include the relationship of competitive balance with the probability a game will sell out. Competitive balance in college football maintains the economic theory from previous research, and highlights the diminishing marginal utility fans receive with less competitive matches.

Key words: Competitive Balance, NCAA, Football, Winning Percentage

Introduction

The NCAA provides game day attendance for all 630 football programs each February. In the 2009-2010 season over 48 million sports fans attended games held during the season. During this period eighteen schools sold out. Of the 18, 2 teams had .500 seasons. Only one of the two reached a bowl game, but averaged sellout crowds. The Big Ten, the third most competitive conference, claimed the top three school attendance spots with Penn State, Ohio State, and Michigan. However, some schools struggled to fill their seats. The Pac 10, the most competitive conference, hosted 60 percent capacity at their stadiums on average[1]. Disparities exist, and simple statistics cannot explain the differences in demand between teams or conferences. How do the fans in the seats react to the quality of teams on the field?

When analyzing the specific data for last year's school some numbers stand out. Winning teams from the mid-majors failed to fill their stadiums. Temple, with multiple winning seasons and bowl appearances averaged 20,515 fans. This is only 29.9 percent of their stadium capacity. BCS conferences had trouble as well. The Pac 10 was the most competitive conference according to the Media Polls and computer rankings. Last year the Pac-10 stadiums sat at 30 percent unoccupied on average. Some teams had rough seasons on the field, but generated large amounts of revenue from the stands. Two teams managed losing records, but were able to average sellout crowds. Those teams were Georgia and Texas with 100,657 and 92,746 fans per game respectively. Unranked Florida also managed to average a sellout crowd of 91,000. Most surprising, Stanford, a Cinderella school with a pro-level quarterback struggled to generate 40,000 fans per game[1].

Inconsistencies in the attendance of sports fans and their demand for tickets have been evaluated by sports economists. One aspect that has drawn economists is the competitive balance

---

[1] CNBC News, "College Football Attendance: By the Numbers," Rovell

between teams. Competitive balance can be described as the distribution of wins in a league, share in championships, or the relative chance related to the outcome of the game. Within the FBS there is roughly an even distribution of wins, although it is not a zero sum game. It is not a zero sum game because some Division IA schools (FBS) play Division IAA schools, but do not compete for the same bowl games. A large determinant of game time attendance is the quality of the teams playing at game time, and the relative uncertainty in the outcome.

Although several sports have been examined through the lens of competitive balance, to date game time measures of competitive balance have not been examined. The most common interpretation explaining why a Stanford hosted 40,000 per game with a 13-1 record and Texas sold out with over 100,000 and a losing record is prestige[1]. It is argued that teams carry prestige, and expectations built upon previous years. While this may be true, could it be possible between games attendance changes to reflect the qualities of the teams playing on any given day? This question poses a significant idea for athletic departments when scheduling games as teams could raise attendance based on the quality of competition.

Literature review

Fort (2003) points out there are two arms of competitive balance to study. The first stream of competitive balance is the statistical study of competitive balance within sports. Traditionally, researchers use indexes or Gini coefficients to explain the distribution of wins, pennants, or championships. The second form, and basis of this paper, is the relationship of attendance, competitive balance, and the affects of business practices. Although not evaluated in this research, business practices include structure changes like free agency or conference realignment[2].

---

[2] The Journal of Sports Economics, "Competitive Balance in Sports Leagues: An Introduction," Fort

Sports economists have measured competitive balance for several sports. Primarily the focus rests on attendance and subsequent revenue. Developed research in Major League Baseball and across European soccer leagues has developed tools to analyze competitive balance. The measures of competitive balance can be calculated over several seasons, a single season, or at the time of the match. Competitive balance can focus on the number of conference championships held by a team, number of pennants held by a baseball club, or the winning percentages of teams. The assumption is that the sport in question is a zero sum sport. Whereas there are only two or three outcomes, win lose or draw (primarily seen in soccer matches) and a set number of opportunities for teams to win. Competitive balance then measures the balance of wins and the relative distribution across a league or sport[2].

Competitive balance can also represent the uncertainty related to the outcome of a game. Studies have used the gambling odds as proxy for the actual competitive balance between teams. While this is not a true index for a distribution it is significant in theory. Odds are created based on the expectations of fans, and develop an approach from inputs such as players, firm spending, and facilities. As a function of the demand for tickets, gambling odds can influence game time attendance. Fans generally have free access to the point spread and free access to sports media in order to make decisions. Fans believe a certain team will win, but this variable also highlights by how many points. The greater the spread, or number of points one team is favored over another, the less competitive the two teams are. The relative chance for the win is diminished for the consumer, and marginal utility drops.

Competitive balance can also be represented by several indexes. Researchers can examine the difference in winning percentages. With a zero a game is perfectly competitive, but a 1 becomes the most uncompetitive. Gini coefficients can be used to illustrate the disparities of titles by team. Some research uses the standard deviation of wins or championships to illustrate the competitive balance across a league.

In a traditional model Krautman et al. (2006) used the standard deviation of winning percentages. Building on prior models, the standard deviation was used in Major League Baseball to determine the likeliness of a team reaching the post season. Data from a Major League Baseball seasons 1950 through 2003 analyzed the effect of competitive balance on dynasties, or teams winning two or more pennants over a five year period. Each team had an opportunity to earn one or more of 44 possible pennants. Diminishing competitive balance between teams in recent years has led to more teams reaching dynasty status. Fans reacted adversely to teams winning multiple dynasties, and attendance dropped for home and away games for the dynasty team. However, Krautman et al. found little change in game time attendance based on their index of competitive balance when compared to their standard deviation measure. They concluded fans react positively to teams capable of making the play offs.[3]

Humphrey (2002) expanded on past research devaluing the Herfindahl-Hirschman Index and the use of standard deviations in winning percentages because little variation was explained by Major League Baseball models. The HHI used winning percentages and not a standard deviation format. A new competitive balance ratio, or CBR, was created by dividing the variation in average win to loss ratio of a team by the variation in average win loss ratio for the host conference in the same year. The CBR is used to calculate team specific variation year to year and league variation year to year. The inter and intra season trends created and used in the same model are significant for studying the long term effects of policy changes as well as the change in attendance based on the variation in competitive balance over a season for individual games. The CBR has been used to explain the impact on new NCAA

[3] Managerial and Decision Economics, "Dynasties versus Pennant Races: Competitive Balance in Major League Baseball." Krautman

regulations on competitive balance.[4] Work after the publication of this article accurately explained the change in competitive balance in NCAA football based on scholarship limits (Sutter and Winkler 2003) and recruiting violations effects on the CBR (Depken and Wilson 2006) across the course of seasons.

Understanding a measure of chance is always present; models then specialize to explain changes during the season. Using past models, Meehan et al. (2007) calculated the competitive balance as the difference in winning percentages between two teams. The scope was to evaluate solely on the game time winning percentages of the two teams. Winning percentages are the easiest measure for fans to gauge the chance associated with individual games. Meehan et al. also checked for asymmetric reactions by the home fans. This is significant because the sign of competitive balance denotes which team is better. A positive index represents a stronger home team winning percentage, negative a stronger visiting team percentage, and a zero represents perfect competition[5]. The initial range of the index was thought to be 0 to 1, explaining only the difference in the two teams as an absolute value. The Meehan et al. index incorporates a -1 lower bound in order to look at games that include a better visiting team. Meehan et al. found asymmetric results where the fans prefer a slightly better home team. Meehan et al. concluded fans prefer a team with .14 to .4 winning percentage, and diminishing returns exist.

Meehan et al. also created reaction variables by multiplying the core measure of competitive balance on additional variables. The researchers found that not only were the competitive balances significant but it changed throughout the season. For instance teams "in the hunt" near the division leader hosted higher attendance with similarly competitive games. This model was stronger yet with the addition of a variable measuring the number of games left in the season. As fewer games remained

[4] Journal of Sports Economics, "Alternative Measures of Competitive Balance in Sports Leagues," Humphrey
[5] Journal of Sports economics, "Competitive Balance and Game Attendance in Major League Baseball," Meehan et al.

more tickets were sold[5]. This contradicts the Krautman et al. (2006) paper and explains why fans attended more games when teams where in the hunt for the post season[3].

Groza (2007) builds on the past research looking at the NCAA Football conference realignment with a different type of competitive balance for NCAA Football. The author assigns the USA Today computer rankings or Sagarin rankings. The rankings include all Division I A and AA schools and take into account winning percentage, conference strength, strength of schedule, and quality wins. The number generated is a probability that team could win 50 percent of the games if the team played all teams at a neutral site. To evaluate competitive balance the difference in ratings (home minus visitor) is squared. The final tobit model yields a significant variable, but only estimates symmetrically the competitive balance because of the squared term. Thus it makes no difference which team is better. As stated in the introduction many argue that college football teams have a prestige level that negates the significance of competitive balance. Groza finds the number of games teams play in the NCAA is not significant as a measure of prestige, but concludes with Price and Sen (2003) that tradition by the number of bowl appearances is the best proxy. Modeling with both bowl appearances and competitive balance Groza still finds the measure of CB significant[6].

This paper builds from previous research and focuses on the NCAA Division IA (FBS) football schedule of 2009 and 2010. This paper develops the Meehan et al. (2007) and Groza (2007) models to specifically look at the affect of competitive balance at game time on attendance. The competitive balance index is developed from the Meehan et al. difference in winning percentages at game time, and a difference in Sagarin predictors by team at game time inspired by the Groza model (2007). This paper builds on past research by evaluating solely competitive balance for college football across years at game time using a panel data set, and estimates if symmetric or asymmetric relationships exist. Lastly,

[6] Managerial Economics, "Conference Call! NCAA Conference Realignment and Football Game Day Attendance," Groza

this paper will account for sold out stadiums, and provide a probability that a team will sell out while including the competitive balance measurements.

The next section develops the theoretical framework from previous models for NCAA College Football. The following section derives empirical models to evaluate competitive balance on attendance, then controls for uncounted demand (sold out stadiums), and finally evaluates the probability a stadium will sell out with competitive balance. The paper concludes with an in interpretation of the results and a conclusion with additional research possibilities.

Theoretical Modeling

The theoretical model is similar to the model developed by Groza (2007). There is only one significant change. The measure of competitive balance takes on two indexes for robustness. Both measures used by Meehan et al. (2007) and Groza (2007) will be evaluated. The purpose is a difference in winning percentages will be a direct index, whereas the Sagarin Predictor will takes on a consumer expectation much like gambling odds.

Bowl appearances were also changed to the standard deviation of bowl appearances for the given number of years. The standard deviation took on the number of appearances, included the number of possible years of attendance in bowl games, and the mean bowl appearances. The standard deviation, mirrored from Humphrey's (2006) model was used in place of a quantitative count for bowl appearances and became significant while maintaining similar parameter estimates for all other variables.

$$\text{Log}\,(\text{Att}_{i,t,g}) = \alpha + \beta_1\text{Game}_{i,t,g} + \beta_2\text{Home}_{i,t,g} + \beta_3\text{Away}_{i,t,g} + \text{CB}_{i,t,g} + e_{i,t,g}$$

The percent change in attendance will be the dependent variable, measuring the percent change from game to game, in year t, for individual home team i. The change in attendance will be measured by specific games posted by the NCAA allowing the qualities of the home and visiting team to affect the overall attendance level. The variables will express consumer tastes and preferences and expectations. The measure of competitive balance will reflect the degree of uncertainty between teams. Consumers prefer for their team to win but do not want to buy a ticket to a blow out. The game time variables of time and day will alter consumer demand based on taste and preferences for later or earlier games, and games played on other days than Saturday.

The price is not accounted for. The NCAA is a price discriminator. Some schools allow students to attend for free, while other teams sell seat licenses and luxury boxes. No detailed price per ticket records exist through NCAA.org. This model must overcome price variation between team that will add to the error term. A team specific fixed effect will account for some of the cross team variation. However, no specific elasticity of demand can be calculated from a percent change in price, for a percent change in attendance.

Empirical Modeling

Initial Modeling will include a pooled panel, two ways fixed effects, OLS model with SAS statistical software. The initial descriptive statistics can be found in appendix II. An F-test run prior with the base model, with competitive balance measure 1, accepted pooled data. Therefore, data from years 2009 and 2010 become a pooled panel.

The base model is as follows:

$$\text{Log}(\text{Att}_{i,t,g}) = \alpha + \lambda \text{CB}_{t,g} + \beta_1 \text{Game}_{i,t,g} + \beta_2 \text{Home}_{i,t,g} + \beta_3 \text{Away}_{i,t,g} + \psi \text{LCB}_{i,t,g} + \phi_t + Y_g + e_{i,t,g}$$

LogATT represents the log of the attendance to game *i* , in year *t*, for game *g*. The interpretation will be a percent change in attendance and will allow for a linear relationship. *CB* represents the competitive balance as an index. Competitive balance is measured in four ways.

1. $\lambda CB_{t,g}$ = Winning_Percentage_Home$_{i,t,g}$ – Winning_ Percentage_Away$_{i,t,g}$
2. $\lambda CB_{t,g}$ =*ABS[* Winning_Percentage_Home$_{i,t,g}$ – Winning_ Percentage_Away$_{i,t,g}$*]*
3. $\lambda CB_{t,g}$ = Sagarin_Predictor_Home$_{i,t,g}$ – Sagarin_Predictor_Away$_{i,t,g}$
4. $\lambda CB_{t,g}$ =*ABS[* Sagarin_Predictor_Home$_{i,t,g}$ – Sagarin_Predictor_Away$_{i,t,g}$*]*

Competitive balance for the above models is measure either as the difference in winning percentages at the time of game for the home and away team (1 and 2), or the difference in the Sagarin Predictor (3 and 4). The first measure is assymetric and allows for a negative number bound between -1 and 1. The second measure is the absolute value and only a spread between the teams without respect to the home team. The third and fourth measures, Sagarins, calculate the gambling odds with respect to the point spread. The third measure accounts for a positive spread if the home team is better, and negative if the away team is better. The results will be asymmetric only if they are significant and different from the fourth measure. The final measure is the absolute value of the spread and gives no respect to the quality of the home team. It is predicted that results will be asymmetric and significant. Theoretically, fans prefer a slightly better home team based on the findings of Meehan et al. (2007)[5]. If the initial OLS modeling finds the absolute value measures insignificant, then the 2nd and 4th measurements of competitive balance will be thrown out.

In the equation above, *Game*, *Home*, and *Away* are a set of variables that control for game and team characteristics. *Game* encompasses three variables. Within the game variable the time, day, and stadium are taken into account. If the game is played in the evening a dummy variable Night takes a value of 1. The predicted sign for Night is either positive or negative as most games are played on Saturday. The second variable, Other Day, is a dummy variable where games not played on Saturday take on a value of 1. The predicted sign is negative or positive because the teams awarded a different day are not randomly selected. Some games take place over holiday weekends on other days than Saturday. Stadium is a dummy variable taking on 1 if the game time attendance is equal to or higher than the listed stadium size. This variable represents excess demand that cannot be calculated. If the stadium reaches or surpasses the capacity it is sold out and will not be calculated as an observation.

Home is a set of variables representative of the home team. *HomeWinPerctPast* includes the three previous years winning percentages. *SDHome* is a vector for program prestige and is calculated as the standard deviation of bowl appearances since 1901. Teams with a higher Standard deviation will have higher prestige. Prestige will raise attendance. *Strength* is a measure of the home team's current winning percentage. The game time winning percentage will not be correlated with the CB measure because it is an index of the two teams. *Home* also includes *conference*. *Conference* is the numbers of championships won buy the group of teams in the same conference. This may be significant in explaining the relative difference in conferences and fan bases. Accounting for idiosyncrasy, unobservable characteristics of the 120 Division IA schools, the variable *TEAM* takes on a specific dummy variable for each school with Wyoming as the base. Wyoming is last in alphabetically, and also the last team to enter the FBS groups of teams. *TEAM* is a descriptive dummy variable that allows for team and university specific fixed effects in order to account for different schools. Further, a time series dummy variable will account for idiosyncrasy between the 2009 and 2010 seasons. The two dummies create a two ways fixed effect OLS model.

Away is a set of three variables encompassing the away team. Initially, while using the base models with all measures of CB bowl appearances and the standard deviations of bowl appearances were evaluated. The *SDAway* vector which represents the program prestige standard deviation of bowl appearances since 1901. The game time winning percentage will not be correlated with the CB measure because it is an index of the two teams. *Away* also includes *conference*. This may be significant in explaining the relative difference in conferences and fan bases.

18 teams sold out during the 2009 and 2010 season. Once a team reaches full capacity the demand function cannot be estimated with equation 1. The model estimates the change in attendance, however changes the true parameter estimates. Model 1 is given an additional variable *Sold_out.* This variable is a binary dummy. Data is sorted to only allow games that are not sold out. After sorting 873 observations of 1305 are left. To evaluate *Sold_Out*, the size of the stadium equals full capacity. If the listed game time attendance is equal or greater, than the variable becomes 1. For the following equation the observation is thrown out.

$$\text{Log}(\text{Att}_{i,t,g}) = \alpha + \lambda \text{CB}_{t,g} + \beta_1\text{Game}_{i,t,g} + \beta_2\text{Home}_{i,t,g} + \beta_3\text{Away}_{i,t,g} + \psi\text{LCB}_{I,t,g} + \phi_t + Y_g + \beta_4\text{Sold\_Out}_{t,g} + e_{i,t,g}$$

Where Sold_Out =0

The final equation evaluates the affect of competitive balance on the probability a stadium sells out. The final model uses the *Sold_out* binary as a dependent variable. The equation estimates the relationship of competitive balance to a stadium reaching full capacity. The equation is posed outside of all previous research to determine if competitive is significant in predicting sellout crowds or just a percent change in attendance.

$$\text{Logit}(\text{Sold\_Out})_{i,t,g} = \alpha + \lambda \text{DiffWinPcnt}_{t,g} + \beta_1\text{Game}_{i,t,g} + \beta_2\text{Home}_{i,t,g} + \beta_3\text{Away}_{i,t,g} + \psi\text{LCB}_{I,t,g} + e_{i,t,g}$$

Data Description

The primary dependent variable is $LogAtt_{i,t,g}$. The attendance is expressed as a percent change in the attendance. Attendance is taken from the NCAA for the 120 Division IA (FBS) schools. Each games attendance is reported to the NCAA, and attendance figures are published in February for the previous season. The first three games are thrown out in order to develop a same-season competitive balance variable similar to both the Meehan et al. (2007) research and the Sagarin statistics (Team and Conference rankings and Predictors)[5]. Games involving Division IAA schools or neutral locations are thrown out as well. Excluding these games allows the fixed effect to account for home team and home fan specific variables like distance. 1137 total games between the 2008-2009 and 2009-2010 seasons are used for the research.

The sold out variable is a measurement of the attendance compared to the stadium capacity. The stadium capacity is found in the NCAA record book. No significant renovations were made during the two seasons observed in this research. The final dependent variable and the dummy variable for the second model is the binary dummy *Sold_Out*. The variable is 1 if the attendance equals or surpasses the stadium capacity. If the attendance falls below the capacity then the sold out variable takes on a value of 0. The Logitistic use of the sold out variable is a binary dependent variable representing the probability a stadium sells out given a set of variables.

Game time variables will include a dummy variable for day in which the game was played. Saturday will be the measure if a game is played on Saturday, resulting in a 0. All other days will be given a 1. The variable will be found on the college football funhouse in the data section including the time of the game. Day will be a second dummy variable. Any game before 7 P.M. Eastern Standard Time will be given a value of 0 and any non-day or night games will be given a 1. The sign may be positive or negative as teams playing on Saturday may not impact the attendance if the game is played during the day or at night. The signs for both day and Saturday could be significant because BCS conference games are often played on Saturday before 7 P.M. Eastern Standard Time. Other times and days are used to highlight mid-major conference games and raise attendance as well as the participating viewing audience.

Away team and home team characteristics will include the stadium size, the number of bowl games the team has played in, and the winning percentages of the three previous seasons. The data will come from the NCAA.org site. The three previous seasons will account for fan expectations. Seasonal results can also be found on the NCAA.org site. In the long run the bowl appearances will display the relative success of the team in the post season as measured by Groza, and similar to Meehan et al. Bowl appearances can be found on each school's athletic department site or through the NCAA records book published annually. Each team belongs to a conference and the relative prestige of the conference will be represented by the number of Collegiate Championships won by member schools.

TEAM is a set of binary dummies coded for each team in order to create a fixed effect for the home team. The base team is Wyoming and accounts for idiosyncrasy, or the unobservable characteristics of the 120 Division IA schools.

Competitive balance is calculated in four equations and is an exogenous variable that encompasses the tastes and preferences of the consumer for tickets in game g, year t, for individual home team i. *CB1* and *CB2* are measures of competitive balance based on the winning percentages of each time at game time. The winning percentages are based on the results for the previous games during the same season. Records are found by team and year on the NCAA.org site. The winning percentage of the away team is subtracted from the home team. The first measure is an index that ranges from -1 to 1. Based on the research of Meehan et al. the sign should be negative when using OLS, with a positive intercept. Based on Meehan et al. (2007) fans prefer a better home team, but interest drops significantly with a rise in competitive balance (teams become less competitive and uncertainty diminishes)[5]. The second measure is symmetric where the fan does not care whether the home or away team is better. The sign is unknown, but the variable should be insignificant.

CB3 and *CB4* are measures of competitive balance based on the Sagarin Predictor from USA Today. Similar to the first two indexes, the signs for *CB3* will be negative if the visiting team is favored. The predictor variable is a proxy for the gambling odds and directly mimics the expectations of fans. Instead of evaluating percentages, the predictor provides point spreads. The home team receives an additional 2.85 points. By subtracting the visiting team from the home team a point spread is created. The index, or spread, predicts a winner and by a number of points. Larger point spreads represent lower balance between teams, but 0 predicts teams will draw (most competitive game). The *CB4* is similar to *CB2.* A change in demand is evaluated for symmetric reactions. Results will explain if fans care if their home team is better than the away team.

The vector of home and away teams is represented by the standard deviation of bowl appearances for all teams since 1901. To allow better description of the variable teams are allowed to keep all bowl and championship appearances without weight to division level. The teams are evaluated on the number of seasons played, for the chances associated with reaching the post season. The mean bowl appearance for 2009 and 2010 is 17.39 and 17.62 respectively. As teams appear in more bowl appearances their prestige rises and the demand for tickets rises as well. The records come from the NCAA records book.

Results

I. OLS

Initially OLS with a fixed effect was used to calculate the percent change in attendance. Results are posted in appendix I. During the initial OLS the models do predict the change in attendance. There is neither change in sign, nor a significant change in the t statistics during this step. The focus is the competitive balance variable. If the competitive balance measure for *CB1* or *CB3* is equal to its respective absolute value (*CB2* and *CB4* respectively), then the reactions of fans is symmetric.

The competitive balance measure, difference in winning percentages, is significant and all others are not. The equations support Meehan et al. (2007). The results are asymmetric. Fans do prefer a better home team to visiting team. The parameter estimate for the 3rd and 4th measure is not only insignificant, but the value is too small to explain any change in attendance. The *winning percentage difference* variable explains that for a 10 percent increase in the home team's winning percentage (while holding the away team's constant) there is a 1.7 percent decline in the game time attendance. Since the competitive balance is an index lowering from the intercept, as the teams become less competitive fans desire to attend the game less, but the confidence interval at the 95 percent levels shows fans prefer the home team to have a 20 to 55 percent better winning percentage.

Games played on other days are significant, while night games are not. This makes sense because all but 4 games outside of Saturdays were smaller conferences. The NCAA traditionally schedules mid-major teams to play on other days in order to allow them media coverage without competing with the larger schools. Attendance drops by a tenth of a percent for any other day than Saturday. Since night games include both mid-majors and BCS schools there is a large variance in the size of attending audiences. Further, since most night games are on Saturday patrons may not have to worry about work on the following day.

The number of total bowl games played by both teams was significant. The home team rose attendance by 1/5 of a percent for every 10 bowl games attended. The visiting team however, only contributed 1/20th of a percent for every 10 bowl games attended. While the parameters are small, but significant, a single percent represents about 1000 people for the 17 largest stadiums according to stadium sizes listed in the NCAA records book.

The winning percentage of the team at the time of the game, and the 3 year winning percentage accounted for larger quantities of fans in the stands. A team with a record of .500 rose attendance by 30 percent. A team with a .500 winning percentage over the past three years could raise attendance by 22 percent. A home team from a conference with 10 NCAA champions saw a significant increase in attendance as well. From 1936 to 2011 the teams of the SEC has won 19 NCAA championships. SEC teams on average have a .6 percent higher attendance rate.

The models have reaffirmed variables found significant in Groza's paper (2007), and the competitive balance indexes support the findings of Meehan et al. (2007). The previous model does not limit observations. There are two possible problems associated with the structure. First there may be some serial correlation related to the modeling procedure. This will be developed in the next draft. The second problem is sold out crowds affect the dependent variable and alters the parameter estimates.

II. Controlling for Sold Out Stadiums

After evaluation of the previous models, only the difference in winning percentages is analyzed as competitive balance. The vectors for home and away bowl appearances are added in place of quantitative bowl appearances. Further, sold out stadiums are thrown from the observations in order to evaluate the change in measurable demand for tickets.

Given the first set of equations the stadium full variable will denote when a stadium reaches full capacity. Since demand cannot be evaluated past the stadium's capacity a binary variable is created from the capacity and attendance. All sold out crowds will be deleted for the next set of equations. Given the findings in the first four equations, and prior research only the true competitive balance indexes will be measured. The absolute values will not be estimated. The sample size observed is 827 games of the original 1035.

Controlling for sold out crowds did not change the previous parameter estimates significantly. The difference in winning percentages originally explained a 10 percent increase in the home winning percentage (while holding the visitor constant) lowered attendance by 10.1 percent. The second model, controlling for sold out stadiums was similar. Given the same scenario, the attendance dropped by 10.7 percent.

III. Probability of a Sell Out

Finally, stadium attendance will be viewed as a Logistic Model. A binary dependent variable, *stadium full,* replaces the Logarithm for attendance. The logistic function will used to determine the affect of competitive balance on the probability a stadium sells out. The Model becomes:

$$\text{Logit(Sold\_Out)}_{i,t,g} = \alpha + \text{CB1}_{i,t,g} + \beta_1\text{Game}_{i,t,g} + \beta_2\text{Home}_{i,t,g} + \beta_3\text{Away}_{i,t,g} + \text{Team} + \varepsilon_{i,t,g}$$

The logistic function through SAS assigns a 1 for the basis of likelihood. Variables then explain the Log-Likelihood that a stadium sells out. The model is significant as highlighted in Appendix III. The competitive balance, difference in winning percentages, is significant at 7 percent. The model suggests that from a 48 percent difference in winning percentage there is a 7 percent decrease in attendance for an additional 10 increase in the winning percentage difference with respect to the home teasm.

In summary the measure is less significant than in traditional models but predicts along the same theory. Fans lose marginal utility as the outcome becomes more certain. Fewer fans attend games when their home team significantly better than the competition.

Conclusions

The models add to previous research by validating the measure of inter and intra seasonal competitive balance for NCAA football at the time of the game. The prestige of the team can accurately account for teams in off seasons. In general fans prefer more competitive matches where the home team is favored, and the prestige through bowl appearances since 1901 adds significantly to the change in attendance.

The results are consistent with Meehan et al. (2007) and the long term measures of CB from Humphrey (2002) when isolating team's prestige as a vector. The only measure of competitive balance that is significant is the difference in winning percentages with respect to home team. The log-likelihood of a stadium selling out is significant, and competitive balance explains that a ten percent rise in the spread between winning percentages lowers attendance by 7.7 percent.

There are limitations to the models. Prices are not included. This negates the worry of endogenous affects, but doesn't portray a true demand function. The demand for tickets most likely changes from inelastic to elastic from more to less prestigious schools. The NCAA is a cartel, where firms can extract more revenue by price discriminating. The price discrimination and market for tickets makes it difficult to develop data for calculation. Further, correlation between program prestige, winning percentage in the past years, and price would most likely be highly correlated.

Also future research should include cohort groups based on model specification. The modeling procedure used in this research assumes that variation for school and conference are constant. However, different schools have different market sizes, student bodies, and level of competition within their conference. Future research can also build from this research by using censored regression. The distribution shows that attendance vastly improves when the home team has at least a winning percent of 30 percent or better. Marginal effects with a tobit could also better highlight the game time determinants affecting changes in attendance.

Works Cited

"Competitive Balance and Conference Realignment: The Case of Big 12 Football | The Sport Journal." Volume 14 | The Sport Journal. Fall 2007. 7 Jan. 2011.

Depken, Craig A., and Dennis P. Wilson. "NCAA Enforcement and Competitive Balance in College Football." *Southern Economic Journal* April 2006 72.4 (2006): 826-45. Web.

Fort, Rodney. ""Competitive Balance in Sports Leagues: An Introduction": Comment." *Journal of Sports Economics*, v. 4 Issue 2, 2003, p. 154-160.

Groza, Mark. "Conference Call! NCAA Conference Realignment and Football Game Day Attendance." Journal of Managerial Economics (2007).

Humphreys, B. R. "Alternative Measures of Competitive Balance in Sports Leagues." Journal of Sports Economics 3.2 (2002): 133-48. Print.

Krautman, Anthony C., and Lawrence Hadley. "Dynasties versus Pennant Races: Competitive Balance in Major League Baseball." Managerial and Decision Economics 27.4 (2006): 287-92. Print.

Meehan, J. W., R. A. Nelson, and T. V. Richardson. "Competitive Balance and Game Attendance in Major League Baseball." Journal of Sports Economics 8.6 (2007): 563-80. Print.

Price DI, Sen KC. 2003. The demand for game day attendance in college football: An analysis of the 1997 Division 1 A season. Managerial and Decision Economics 2: 35–46.

Rovell, Darren. "College Football Attendance: By The Numbers." *CNBC Mobile Home*. CNBC, 13 Jan. 2011. Web. 1 Apr. 2011. <http://www.cnbc.com/id/41060717>.

Schmidt, M. B., and D. J. Berri. "Competitive Balance and Attendance: The Case of Major League Baseball." Journal of Sports Economics 2.2 (2001): 145-67. Print.

Sutter, Daniel, and Stephen Winkler. "NCAA Scholarship Limits and Competitive Balance in College Football." *Journal of Sports Economics* February 4.1 (2003): 3-18. Print.

*NCAA.org/stats/football (attendance/game date time and results)

*USA Today Sagarin Rankings (college Football by Year teams and conferences)

*NCAA.org/records book (bowl appearances, basic stats and facts)

Appendix I. Tables

Table I. The pooled panel OLS is used for all measures using the bowl appearances in a quantitative method.

Variable	Expected Sign	Model 1 Parameter	Model 2 Prameter	Model 3 Parameter	Model 4 Parameter
Intercept	+	9.664	9.307		9.287
CB1	-	-.1017*** (-3.10)			
CB2 (ABS)	-		.0006 (1.47)		
CB3	-			-.00111 (-1.32)	
CB4 (ABS)	-				.0020* (1.94)
Night	+,-	.0805*** (3.20)	.0728*** (3.17)	.1591*** (2.39)	.0834 (3.30)
Other Day	+,-	-.1354*** (-4.26)	-.1014*** (-4.18)	-.1295*** (-4.08)	-.1284*** (-4.06)
Home bowl appearance	+	-.0093* (-1.69)	-.0090 (-1.65)	-.0094* (-1.70)	-.0089 (-1.61)
Winning percent (H)	+	.3088*** (4.99)	.2399*** (4.54)	.2354*** (4.17)	.2261*** (4.04)
(H)Long Run Win Percent	+	.1521** (2.30)	.1645*** (2.32)	.1591*** (2.39)	..1721*** (2.59)
(H) Conf. Champions	+	-.05128 (-.44)	-.0046 (-.33)	-.0332 (-.28)	-.0559 (-.47)
Visit bowl	+	.0030*** (3.51)	.0034*** (3.37)	.0031*** (3.29)	.0039*** (4.69)
R Sqr		.9129	.9064	.9119	.9121
RMSE		.21334	.3487	.2154	.2142
F Value		61.03	60.74	60.30	60.49

Notes 99% Confidence Level ***, 95% Confidence Level **, 90% Confidence Level *

Number of observations: 1305

F.E. TS=Year, CS=Team

Table II. The pooled panel OLS is used for all measures using the bowl appearances in a quantitative method. Only Competitive Balance in the form of difference in winning percentages is used. Table I shows asymmetric results, where only CB1 is significant (all others are considered zero) and home and away vectors are substituted for bowl appearances. The vectors do not interact. The variable SDHome and SDAway are standard deviations of bowl appearances.

Variable	Expected Sign	Model CB1 Parameter
Intercept	+	9.694
WinDiff	-	-.10073*** (-3.05)
Night	+,-	.08029*** (3.18)
Other Day	+,-	-.1344*** (-4.26)
SDHome	+	-6.5751* (-1.78)
Winning percent (H)	+	.31532*** (2.31)
(H)Long Run Win Percent	+	.1521** (2.30)
(H) Conf. Champions	+	-.04907 (-.42)
SDAway	+	2.145*** 3.31
R Sqr		.9127
RMSE		.2135
F Value		61.95

Notes 99% Confidence Level ***, 95% Confidence Level **, 90% Confidence Level *

Number of observations: 1305

F.E. TS=Year, CS=Team

Appendix II. SAS Code

```
data work.football;
  set football;

/*Dummy variable for unrepresented demand, thus any attendance figure below the capacity by one is considered
not sold out*/
  sold_out=.;
    if attendance > (not_sold_out) then sold_out= 1;
    else sold_out=0;

/*TESTING PURPOSES */
        lncb3=log(cb3);

        lncb4= log(cb4);

        yeardummy=.;
        if year=2009 then yeardummy=0;
        else yeardummy=1;

/*  Log Variables   */

  lnatt=log(attendance);


/*F.E. Variables for Schools/Teams; Binary Dummies with Wyoming University as the base school
because newest entrant to FBS and last by a to z.  There are 120 Division 1A or FBS schools and
119 dummies*/

  af=.;
    if home_team_code=1 then af=1;
    else af=0;

  ak=.;
    if home_team_code=2 then ak=1;
    else ak=0;

  bama=.;
    if home_team_code=3 then bama=1;
    else bama=0;

  az=.;
    if home_team_code=4 then az=1;
    else az=0;

  asu=.;
    if home_team_code=5 then asu=1;
    else asu=0;

  ark=.;
    if home_team_code=6 then ark=1;
    else ark=0;

  arks=.;
    if home_team_code=7 then arks=1;
    else arks=0;
```

```
arm=.;
  if home_team_code=8 then arm=1;
  else arm=0;

aub=.;
  if home_team_code=9 then aub=1;
  else aub=0;

ball=.;
  if home_team_code=10 then ball=1;
  else ball=0;

bay=.;
  if home_team_code=11 then bay=1;
  else bay=0;

boi=.;
  if home_team_code=12 then boi=1;
  else boi=0;

bc=.;
  if home_team_code=13 then bc=1;
  else bc=0;

bg=.;
  if home_team_code=14 then bg=1;
  else bg=0;

buff=.;
  if home_team_code=15 then buff=1;
  else buff=0;

byu1=.;
  if home_team_code=16 then byu1=1;
  else byu1=0;

cal=.;
  if home_team_code=17 then cal=1;
  else cal=0;

cen=.;
  if home_team_code=18 then cen=1;
  else cen=0;

cin=.;
  if home_team_code=19 then cin=1;
  else cin=0;

cle=.;
  if home_team_code=20 then cle=1;
  else cle=0;

col=.;
  if home_team_code=21 then col=1;
  else col=0;

cols=.;
  if home_team_code=22 then cols=1;
  else cols=0;

ucon=.;
  if home_team_code=23 then ucon=1;
  else ucon=0;

duke1=.;
  if home_team_code=24 then duke1=1;
```

```
   else duke1=0;

ecar=.;
   if home_team_code=25 then ecar=1;
   else ecar=0;

emic=.;
   if home_team_code=26 then emic=1;
   else emic=0;

flu=.;
   if home_team_code=27 then flu=1;
   else flu=0;

fatl=.;
   if home_team_code=28 then fatl=1;
   else fatl=0;

flo=.;
   if home_team_code=29 then flo=1;
   else flo=0;

fsu=.;
   if home_team_code=30 then fsu=1;
   else fsu=0;

fres=.;
   if home_team_code=31 then fres=1;
   else fres=0;

geo=.;
   if home_team_code=32 then geo=1;
   else geo=0;

gt=.;
   if home_team_code=33 then gt=1;
   else gt=0;

hi=.;
   if home_team_code=34 then hi=1;
   else hi=0;

hou=.;
   if home_team_code=35 then hou=1;
   else hou=0;

idho=.;
   if home_team_code=36 then idho=1;
   else idho=0;

ill=.;
   if home_team_code=37 then ill=1;
   else ill=0;

ind=.;
   if home_team_code=38 then ind=1;
   else ind=0;

iow=.;
   if home_team_code=39 then iow=1;
   else iow=0;

isu=.;
   if home_team_code=40 then isu=1;
   else isu=0;
```

```
kan=.;
  if home_team_code=41 then kan=1;
  else kan=0;

ksu=.;
  if home_team_code=42 then ksu=1;
  else ksu=0;

kent=.;
  if home_team_code=43 then kent=1;
  else kent=0;

ky=.;
  if home_team_code=44 then ky=1;
  else ky=0;

lal=.;
  if home_team_code=45 then lal=1;
  else lal=0;

lam=.;
  if home_team_code=46 then lam=1;
  else lam=0;

lat=.;
  if home_team_code=47 then lat=1;
  else lat=0;

lsu1=.;
  if home_team_code=49 then lsu1=1;
  else lsu1=0;

lou=.;
  if home_team_code=48 then lou=1;
  else lou=0;

mar=.;
  if home_team_code=50 then mar=1;
  else mar=0;

mary=.;
  if home_team_code=51 then mary=1;
  else mary=0;

mem=.;
  if home_team_code=52 then mem=1;
  else mem=0;

miaf=.;
  if home_team_code=53 then miaf=1;
  else miaf=0;

miao=.;
  if home_team_code=54 then miao=1;
  else miao=0;

ugly=.;
  if home_team_code=55 then ugly=1;
  else ugly=0;

msu=.;
  if home_team_code=56 then msu=1;
  else msu=0;

midt=.;
```

```
    if home_team_code=57 then midt=1;
    else midt=0;

  minn=.;
    if home_team_code=58 then minn=1;
    else minn=0;

  omiss=.;
    if home_team_code=59 then omiss=1;
    else omiss=0;

  miss=.;
    if home_team_code=60 then miss=1;
    else miss=0;

  miz=.;
    if home_team_code=61 then miz=1;
    else miz=0;

  bad=.;
    if home_team_code=62 then bad=1;
    else bad=0;

  neb=.;
    if home_team_code=63 then neb=1;
    else neb=0;

  nev=.;
    if home_team_code=64 then nev=1;
    else nev=0;

  newm=.;
    if home_team_code=65 then newm=1;
    else newm=0;

  nmsu=.;
    if home_team_code=66 then nmsu=1;
    else nmsu=0;

  nc=.;
    if home_team_code=67 then nc=1;
    else nc=0;

  ncs=.;
    if home_team_code=68 then ncs=1;
    else ncs=0;

  ntex=.;
    if home_team_code=69 then ntex=1;
    else ntex=0;

  nill=.;
    if home_team_code=70 then nill=1;
    else nill=0;

  nor=.;
    if home_team_code=71 then nor=1;
    else nor=0;

  dame=.;
    if home_team_code=72 then dame=1;
    else dame=0;

  ohiou=.;
    if home_team_code=73 then ohiou=1;
    else ohiou=0;
```

```
the_ohio_state_university=.;
  if home_team_code=74 then the_ohio_state_university=1;
  else the_ohio_state_university=0;

ok=.;
  if home_team_code=75 then ok=1;
  else ok=0;

oks=.;
  if home_team_code=76 then oks=1;
  else oks=0;

or=.;
  if home_team_code=77 then or=1;
  else or=0;

ors=.;
  if home_team_code=78 then ors=1;
  else ors=0;

psu=.;
  if home_team_code=79 then psu=1;
  else psu=0;

pitt=.;
  if home_team_code=80 then pitt=1;
  else pitt=0;

pur=.;
  if home_team_code=81 then pur=1;
  else pur=0;

ric=.;
  if home_team_code=82 then ric=1;
  else ric=0;

rut=.;
  if home_team_code=83 then rut=1;
  else rut=0;

sds=.;
  if home_team_code=84 then sds=1;
  else sds=0;

sjs=.;
  if home_team_code=85 then sjs=1;
  else sjs=0;

smu1=.;
  if home_team_code=86 then smu1=1;
  else smu1=0;

scar=.;
  if home_team_code=87 then scar=1;
  else scar=0;

sflo=.;
  if home_team_code=88 then sflo=1;
  else sflo=0;

usc1=.;
  if home_team_code=89 then usc1=1;
  else usc1=0;

smiss=.;
```

```
  if home_team_code=90 then smiss=1;
  else smiss=0;

stan=.;
  if home_team_code=91 then stan=1;
  else stan=0;

syr=.;
  if home_team_code=92 then syr=1;
  else syr=0;

tcu1=.;
  if home_team_code=93 then tcu1=1;
  else tcu1=0;

temp=.;
  if home_team_code=94 then temp=1;
  else temp=0;

tenn=.;
  if home_team_code=95 then tenn=1;
  else tenn=0;

tex=.;
  if home_team_code=96 then tex=1;
  else tex=0;

tam=.;
  if home_team_code=97 then tam=1;
  else tam=0;

ttech=.;
  if home_team_code=98 then ttech=1;
  else ttech=0;

tole=.;
  if home_team_code=99 then tole=1;
  else tole=0;

troy1=.;
  if home_team_code=100 then troy1=1;
  else troy1=0;

tul=.;
  if home_team_code=101 then tul=1;
  else tul=0;

tuls=.;
  if home_team_code=102 then tuls=1;
  else tuls=0;

uab1=.;
  if home_team_code=103 then uab1=1;
  else uab1=0;

ucf1=.;
  if home_team_code=104 then ucf1=1;
  else ucf1=0;

ucla1=.;
  if home_team_code=105 then ucla1=1;
  else ucla1=0;

unlv=.;
  if home_team_code=106 then unlv=1;
  else unlv=0;
```

```
utah1=.;
  if home_team_code=107 then utah1=1;
  else utah1=0;

utahs=.;
  if home_team_code=108 then utahs=1;
  else utahs=0;

utep1=.;
  if home_team_code=109 then utep1=1;
  else utep1=0;

van=.;
  if home_team_code=110 then van=1;
  else van=0;

vir=.;
  if home_team_code=111 then vir=1;
  else vir=0;

vat=.;
  if home_team_code=112 then vat=1;
  else vat=0;

wake=.;
  if home_team_code=113 then wake=1;
  else wake=0;

was=.;
  if home_team_code=114 then was=1;
  else was=0;

wsu=.;
  if home_team_code=115 then wsu=1;
  else wsu=0;

wva=.;
  if home_team_code=116 then wva=1;
  else wva=0;

wky=.;
  if home_team_code=117 then wky=1;
  else wky=0;

wmich=.;
  if home_team_code=118 then wmich=1;
  else wmich=0;

wisc=.;
  if home_team_code=119 then wisc=1;
  else wisc=0;

        run;

proc corr;
var home_bowl_app conf_champ h_sag;
run;

proc means
  N Mean Median STD min max clm
  maxdec=2
  data= work.football;
```

```
  var sold_out h_sag a_sag home_bowl_app away_bowl_app win_perct_past sold_out attendance home_perct
visit_perct other_day night cb1 cb2 cb3 cb4 stdhome stdaway difstd;
  run;

proc univariate data = football noprint;
 histogram lnatt*cb1 / vscale = count;
run;

proc sort;
by sold_out;

/* PART I */
/* cb1 is the difference in winning percentages with respect to the home team, estimated at .xx levels, where cb1
ranges from -1 to 1*/
proc reg;
 model  lnatt=cb1 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ
yeardummy af ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar
emic flu fatl flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf
miao ugly msu midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou
the_ohio_state_university ok oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp
tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky
wmich wisc/acov;
where sold_out=0;
run;
/* cb2 is the absolute value of the difference in winning percentages, estimated at .xx levels, where cb2 ranges
from 0 to 1*/
proc reg;
 model  lnatt=cb2 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ af
ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar emic flu fatl
flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf miao ugly msu
midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou the_ohio_state_university ok
oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp tenn tex tam ttech tole troy1 tul
tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky wmich wisc yeardummy/acov;
where sold_out=0;
run;
/* where cb3 is the point spread or gambling odds with respect to the home team and can go negative if the away
team is better*/
proc reg;
 model  lnatt=cb3 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ
yeardummy af ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar
emic flu fatl flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf
miao ugly msu midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou
the_ohio_state_university ok oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp
tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky
wmich wisc/acov;
where sold_out=0;
run;
/* where lncb3 will illustrate the percentage of points favored on the percent change in attendance*/
proc reg;
 model  lnatt=lncb3 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ
yeardummy af ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar
emic flu fatl flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf
miao ugly msu midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou
```

```
the_ohio_state_university ok oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp
tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky
wmich wisc/acov;
where sold_out=0;
run;
/* cb4 is the absolute point advantage without respect to the home team, unbiased estimator of point spread or
gambling odds */
proc reg;
 model   lnatt=cb4 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ
yeardummy af ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar
emic flu fatl flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf
miao ugly msu midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou
the_ohio_state_university ok oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp
tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky
wmich wisc/acov;
where sold_out=0;
run;
/* where lncb3 will illustrate the percentage of points favored on the percent change in attendance*/
proc reg;
 model   lnatt=lncb4 home_bowl_app win_perct_past night other_day away_bowl_app home_perct conf_champ
yeardummy af ak bama az asu ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar
emic flu fatl flo fsu fres geo gt hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf
miao ugly msu midt minn omiss miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou
the_ohio_state_university ok oks or ors psu pitt pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp
tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1 unlv utah1 utahs utep1 van vir vat wake was wsu wva wky
wmich wisc/acov;
where sold_out=0;
run;

/*  PART II  */

proc reg;
 model   lnatt=cb1 win_perct_past night other_day home_perct conf_champ stdhome stdaway af ak bama az asu
ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar emic flu fatl flo fsu fres geo gt
hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf miao ugly msu midt minn omiss
miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou the_ohio_state_university ok oks or ors psu pitt
pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1
unlv utah1 utahs utep1 van vir vat wake was wsu wva wky wmich wisc yeardummy/acov ;
        where sold_out=0;
run;

proc reg;
 model   lnatt=cb3 win_perct_past night other_day home_perct conf_champ stdhome stdaway af ak bama az asu
ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar emic flu fatl flo fsu fres geo gt
hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf miao ugly msu midt minn omiss
miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou the_ohio_state_university ok oks or ors psu pitt
pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1
unlv utah1 utahs utep1 van vir vat wake was wsu wva wky wmich wisc yeardummy/acov;
        where sold_out=0;
run;

proc reg;
```

```
model   lnatt=lncb3 win_perct_past night other_day home_perct conf_champ stdhome stdaway af ak bama az asu
ark arks arm aub ball bay boi bc bg buff byu1 cal cen cin cle col cols ucon duke1 ecar emic flu fatl flo fsu fres geo gt
hi hou idho ill ind iow isu kan ksu kent ky lal lam lat lsu1 lou mar mary mem miaf miao ugly msu midt minn omiss
miss miz bad neb nev newm nmsu nc ncs ntex nill nor dame ohiou the_ohio_state_university ok oks or ors psu pitt
pur ric rut sds sjs smu1 scar sflo usc1 smiss stan syr tcu1 temp tenn tex tam ttech tole troy1 tul tuls uab1 ucf1 ucla1
unlv utah1 utahs utep1 van vir vat wake was wsu wva wky wmich wisc yeardummy/acov;
        where sold_out=0;
run;


/*  PART III */
ods graphics on;
proc logistic data=football descending;
  model sold_out = cb1 win_perct_past night other_day home_perct conf_champ stdhome stdaway;
;
run;
ods graphics off;

proc logistic data= football descending;

  model sold_out = cb3 win_perct_past night other_day home_perct conf_champ stdhome stdaway;

run;
```

Appendix III: Results

The SAS System

The MEANS Procedure

Variable	Label	N	Mean	Median	Std Dev	Minimum	Maximum	Lower 95% CL for Mean	Upper 95% CL for Mean
sold_out		1035	0.20	0.00	0.40	0.00	1.00	0.18	0.23
h_sag	h_sag	1035	70.58	70.42	11.99	42.41	100.91	69.85	71.32
a_sag	a_sag	1035	69.79	69.37	12.11	42.41	100.91	69.05	70.52
home_bowl_app	home_bowl_app	1035	17.54	14.00	13.60	0.00	58.00	16.71	18.37
away_bowl_app	away_bowl_app	1035	16.40	13.00	13.31	0.00	58.00	15.58	17.21
win_perct_past	win_perct_past	1035	0.52	0.54	0.22	0.00	1.00	0.51	0.53
Attendance	Attendance	1035	44027.40	40643.00	26392.13	1404.00	113065.00	42417.65	45637.16
home_perct	home_perct	1035	0.54	0.56	0.26	0.00	1.00	0.53	0.56
visit_perct	visit_perct	1035	0.54	0.56	0.27	0.00	1.00	0.53	0.56
other_day	other_day	1035	0.10	0.00	0.30	0.00	1.00	0.08	0.12
night	night	1035	0.22	0.00	0.41	0.00	1.00	0.19	0.24
cb1	cb1	1035	-0.00	0.00	0.34	-1.00	1.00	-0.02	0.02
cb2	cb2	1035	0.27	0.25	0.20	0.00	1.00	0.26	0.28
cb3	cb3	1035	3.81	3.29	14.59	-46.57	48.71	2.92	4.70
cb4	cb4	1035	12.02	10.13	9.09	0.00	48.71	11.47	12.58
stdhome	stdhome	1035	0.03	0.03	0.02	0.00	0.06	0.03	0.03
stdaway	stdaway	1035	0.03	0.03	0.02	0.00	0.06	0.03	0.03
difstd	difstd	1035	0.00	0.00	0.02	-0.05	0.06	0.00	0.00

The CORR Procedure

3 Variables: home_bowl_app conf_champ h_sag

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
home_bowl_app	1035	17.53623	13.59519	18150	0	58.00000
conf_champ	1035	7.61449	7.34868	7881	0	19.00000
h_sag	1035	70.58359	11.99382	73054	42.41000	100.91000

Simple Statistics

Variable	Label
home_bowl_app	home_bowl_app
conf_champ	conf_champ
h_sag	h_sag

Pearson Correlation Coefficients, N = 1035
Prob > |r| under H0: Rho=0

	home_ bowl_app	conf_ champ	h_sag
home_bowl_app	1.00000	0.64104	0.66018
home_bowl_app		<.0001	<.0001
conf_champ	0.64104	1.00000	0.49893
conf_champ	<.0001		<.0001
h_sag	0.66018	0.49893	1.00000
h_sag	<.0001	<.0001	

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	827

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	336.13822	2.77800	61.03	<.0001
Error	705	32.08883	0.04552		
Corrected Total	826	368.22705			

Root MSE	0.21334	R-Square	0.9129
Dependent Mean	10.32083	Adj R-Sq	0.8979
Coeff Var	2.06713		

The REG Procedure
Model: MODEL cb1_a
Dependent Variable: lnatt

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9.66834	0.24728	39.10	<.0001
cb1	cb1	1	-0.10169	0.03284	-3.10	0.0020
home_bowl_app	home_bowl_app	1	-0.00933	0.00550	-1.69	0.0906
win_perct_past	win_perct_past	1	0.15208	0.06618	2.30	0.0219
night	night	1	0.08054	0.02519	3.20	0.0014
other_day	other_day	1	-0.13450	0.03156	-4.26	<.0001
away_bowl_app	away_bowl_app	1	0.00303	0.00086306	3.51	0.0005
home_perct	home_perct	1	0.30877	0.06187	4.99	<.0001
conf_champ	conf_champ	1	-0.05128	0.11695	-0.44	0.6612
yeardummy		1	0.01828	0.01562	1.17	0.2422

The REG Procedure
Model: Model cb2_a
Dependent Variable: lnatt

Heteroscedasticity Consistent Covariance of Estimates

Variable	Label	Intercept	cb2	home_bowl_app	win_perct_past
Intercept	Intercept	0.0187406575	-0.000974506	-0.000167679	-0.002136922
cb2	cb2	-0.000974506	0.001596194	-0.000047174	0.0002536099
home_bowl_app	home_bowl_app	-0.000167679	-0.000047174	0.0000468766	-0.000038144
win_perct_past	win_perct_past	-0.002136922	0.0002536099	-0.000038144	0.0037116572
night	night	0.000045032	-0.000018721	0.0000112757	-0.000147777
other_day	other_day	-0.000072597	0.0001721525	-0.000027288	0.0003439635
away_bowl_app	away_bowl_app	-9.900222E-6	9.9835528E-7	-1.24124E-7	-4.080196E-6
home_perct	home_perct	-0.00007689	0.0000155639	-0.00001093	0.0005112721
conf_champ	conf_champ	-0.006792405	0.000268345	-0.00001533	0.0001741972

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	827

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	335.78166	2.77506	60.30	<.0001
Error	705	32.44538	0.04602		
Corrected Total	826	368.22705			

Root MSE	0.21453	R-Square	0.9119
Dependent Mean	10.32083	Adj R-Sq	0.8968
Coeff Var	2.07858		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9.67309	0.24894	38.86	<.0001
cb3	cb3	1	-0.00111	0.00084353	-1.32	0.1884
home_bowl_app	home_bowl_app	1	-0.00938	0.00553	-1.70	0.0903
win_perct_past	win_perct_past	1	0.15911	0.06652	2.39	0.0170
night	night	1	0.08043	0.02540	3.17	0.0016
other_day	other_day	1	-0.12945	0.03170	-4.08	<.0001
away_bowl_app	away_bowl_app	1	0.00314	0.00095449	3.29	0.0011
home_perct	home_perct	1	0.23538	0.05651	4.17	<.0001
conf_champ	conf_champ	1	-0.03316	0.11770	-0.28	0.7783
yeardummy		1	0.01591	0.01569	1.01	0.3111

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	484
Number of Observations with Missing Values	343

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	115	177.40422	1.54265	39.01	<.0001
Error	368	14.55352	0.03955		
Corrected Total	483	191.95774			

Root MSE	0.19887	R-Square	0.9242
Dependent Mean	10.41258	Adj R-Sq	0.9005
Coeff Var	1.90986		

The REG Procedure
Model: lncb3
Dependent Variable: lnatt

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9.61091	0.26149	36.75	<.0001
lncb3		1	0.00007651	0.01047	0.01	0.9942
home_bowl_app	home_bowl_app	1	-0.00321	0.00593	-0.54	0.5881
win_perct_past	win_perct_past	1	0.22376	0.10229	2.19	0.0293
night	night	1	0.08192	0.03350	2.45	0.0149
other_day	other_day	1	-0.10535	0.04212	-2.50	0.0128
away_bowl_app	away_bowl_app	1	0.00115	0.00131	0.88	0.3821
home_perct	home_perct	1	0.17347	0.07384	2.35	0.0193
conf_champ	conf_champ	1	-0.02217	0.11342	-0.20	0.8451
yeardummy		1	0.00755	0.02133	0.35	0.7237

The REG Procedure
Model: cb4
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	827

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	335.87418	2.77582	60.49	<.0001
Error	705	32.35287	0.04589		
Corrected Total	826	368.22705			

Root MSE	0.21422	R-Square	0.9121
Dependent Mean	10.32083	Adj R-Sq	0.8971
Coeff Var	2.07562		

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9.68475	0.24820	39.02	<.0001
cb4	cb4	1	0.00196	0.00101	1.94	0.0531
home_bowl_app	home_bowl_app	1	-0.00889	0.00553	-1.61	0.1084
win_perct_past	win_perct_past	1	0.17214	0.06652	2.59	0.0099
night	night	1	0.08336	0.02528	3.30	0.0010
other_day	other_day	1	-0.12840	0.03161	-4.06	<.0001
away_bowl_app	away_bowl_app	1	0.00394	0.00084129	4.69	<.0001
home_perct	home_perct	1	0.22607	0.05590	4.04	<.0001
conf_champ	conf_champ	1	-0.05586	0.11763	-0.47	0.6350
yeardummy		1	0.01064	0.01578	0.67	0.5003

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	826
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	335.70441	2.77442	60.29	<.0001
Error	704	32.39817	0.04602		
Corrected Total	825	368.10258			

Root MSE	0.21452	R-Square	0.9120
Dependent Mean	10.32125	Adj R-Sq	0.8969
Coeff Var	2.07846		

Parameter Estimates

Variable	Label	DF	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.13580	71.40	<.0001
lncb4		1	0.00727	0.07	0.9427
home_bowl_app	home_bowl_app	1	0.00683	-1.39	0.1665
win_perct_past	win_perct_past	1	0.06192	2.57	0.0105
night	night	1	0.02426	3.44	0.0006
other_day	other_day	1	0.03750	-3.38	0.0008
away_bowl_app	away_bowl_app	1	0.00062864	5.80	<.0001
home_perct	home_perct	1	0.05034	4.41	<.0001
conf_champ	conf_champ	1	0.05916	-0.70	0.4833
yeardummy		1	0.01539	1.01	0.3115

The REG Procedure
Model: cb1_controlled_sold_out
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	827

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	336.09714	2.77766	60.95	<.0001
Error	705	32.12990	0.04557		
Corrected Total	826	368.22705			

Root MSE	0.21348	R-Square	0.9127
Dependent Mean	10.32083	Adj R-Sq	0.8978
Coeff Var	2.06845		

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9.69411	0.25169	38.52	<.0001
cb1	cb1	1	-0.10073	0.03298	-3.05	0.0023
win_perct_past	win_perct_past	1	0.15317	0.06641	2.31	0.0214
night	night	1	0.08029	0.02522	3.18	0.0015
other_day	other_day	1	-0.13442	0.03159	-4.26	<.0001
home_perct	home_perct	1	0.30445	0.06208	4.90	<.0001
conf_champ	conf_champ	1	-0.04907	0.11700	-0.42	0.6750
stdhome	stdhome	1	-6.57506	3.69809	-1.78	0.0758
stdaway	stdaway	1	2.14516	0.64881	3.31	0.0010

The REG Procedure
Model: cb3_contorlled_sold_out
Dependent Variable: lnatt

Number of Observations Read	827
Number of Observations Used	827

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	335.75575	2.77484	60.25	<.0001
Error	705	32.47129	0.04606		
Corrected Total	826	368.22705			

Root MSE	0.21461	R-Square	0.9118
Dependent Mean	10.32083	Adj R-Sq	0.8967
Coeff Var	2.07941		

The SAS

Parameter Estimates

Variable	Label	DF	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.16047	60.44	<.0001
cb3	cb3	1	0.00076700	-1.49	0.1372
win_perct_past	win_perct_past	1	0.06229	2.57	0.0104
night	night	1	0.02433	3.29	0.0011
other_day	other_day	1	0.03745	-3.46	0.0006
home_perct	home_perct	1	0.05144	4.51	<.0001
conf_champ	conf_champ	1	0.06024	-0.51	0.6120
stdhome	stdhome	1	6.44126	-1.05	0.2931
stdaway	stdaway	1	0.60959	3.65	0.0003

The REG Procedure
Model: MODEL1
Dependent Variable: lnatt

Parameter Estimates

			---Heteroscedasticity Consistent--		
Variable	Label	DF	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.16133	59.52	<.0001
lncb3		1	0.00804	-0.06	0.9542
win_perct_past	win_perct_past	1	0.10439	2.14	0.0328
night	night	1	0.02894	2.85	0.0047
other_day	other_day	1	0.03717	-2.82	0.0051
home_perct	home_perct	1	0.07200	2.38	0.0179
conf_champ	conf_champ	1	0.05783	-0.36	0.7185
stdhome	stdhome	1	3.43606	-0.18	0.8591
stdaway	stdaway	1	0.73678	0.72	0.4696

The LOGISTIC Procedure

Model Information

Data Set	WORK.FOOTBALL
Response Variable	sold_out
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1035
Number of Observations Used	1035

Response Profile

Ordered Value	sold_out	Total Frequency
1	1	208
2	0	827

Probability modeled is sold_out=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1040.600	772.026
SC	1045.542	816.506
-2 Log L	1038.600	754.026

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	284.5733	8	<.0001
Score	238.5848	8	<.0001
Wald	166.0852	8	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.7931	0.4962	187.4449	<.0001
cb1	1	-0.7348	0.4007	3.3634	0.0667
win_perct_past	1	2.5443	0.5472	21.6196	<.0001
night	1	0.0704	0.2766	0.0648	0.7991
other_day	1	-0.1253	0.4053	0.0957	0.7571
home_perct	1	3.0232	0.5958	25.7492	<.0001
conf_champ	1	0.0768	0.0171	20.0892	<.0001
stdhome	1	12.5850	8.1124	2.4066	0.1208
stdaway	1	27.6928	6.7379	16.8922	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
cb1	0.480	0.219	1.052
win_perct_past	12.734	4.357	37.217
night	1.073	0.624	1.845
other_day	0.882	0.399	1.952
home_perct	20.557	6.395	66.083
conf_champ	1.080	1.044	1.117
stdhome	>999.999	0.036	>999.999
stdaway	>999.999	>999.999	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.3	Somers' D	0.687
Percent Discordant	15.5	Gamma	0.689
Percent Tied	0.2	Tau-a	0.221
Pairs	172016	c	0.844

The LOGISTIC Procedure

Model Information

Data Set	WORK.FOOTBALL
Response Variable	sold_out
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1035
Number of Observations Used	1035

Response Profile

Ordered Value	sold_out	Total Frequency
1	1	208
2	0	827

Probability modeled is sold_out=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1040.600	775.406
SC	1045.542	819.886
-2 Log L	1038.600	757.406

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	281.1933	8	<.0001
Score	235.7001	8	<.0001
Wald	164.6042	8	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.5649	0.4777	188.8411	<.0001
cb3	1	-0.00086	0.00859	0.0099	0.9206
win_perct_past	1	2.5853	0.5500	22.0939	<.0001
night	1	0.0713	0.2761	0.0667	0.7961
other_day	1	-0.0615	0.4024	0.0233	0.8786
home_perct	1	2.3277	0.4905	22.5234	<.0001
conf_champ	1	0.0777	0.0171	20.6488	<.0001
stdhome	1	12.8374	8.2961	2.3944	0.1218
stdaway	1	31.9355	7.2894	19.1942	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
cb3	0.999	0.982	1.016
win_perct_past	13.267	4.514	38.990
night	1.074	0.625	1.845
other_day	0.940	0.427	2.069
home_perct	10.255	3.921	26.817
conf_champ	1.081	1.045	1.118
stdhome	>999.999	0.033	>999.999
stdaway	>999.999	>999.999	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.1	Somers' D	0.683
Percent Discordant	15.7	Gamma	0.685
Percent Tied	0.2	Tau-a	0.220
Pairs	172016	c	0.842