DISCUSSION PAPER SERIES

No. 5239

A CONSISTENT WEIGHTED RANKING SCHEME WITH AN APPLICATION TO NCAA COLLEGE FOOTBALL RANKINGS

Itay Fainmesser, Chaim Fershtman and Neil Gandal

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP5239.asp

ISSN 0265-8003

A CONSISTENT WEIGHTED RANKING SCHEME WITH AN APPLICATION TO NCAA COLLEGE FOOTBALL RANKINGS

Itay Fainmesser, Tel Aviv University
Chaim Fershtman, Tel Aviv University, Erasmus University Rotterdam and CEPR
Neil Gandal, Tel Aviv University and CEPR

Discussion Paper No. 5239
September 2005

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5239

September 2005

ABSTRACT

A Consistent Weighted Ranking Scheme with an Application to NCAA College Football Rankings

The NCAA college football ratings, in which the "so-called" national champion is determined, has been plagued by controversies the last few years. The difficulty arises because there is a need to make a complete ranking of teams even though each team has a different schedule of games with a different set of opponents. A similar problem arises whenever one wants to establish a ranking of patents or academic journals, etc. in which the raw data are (incomplete) bilateral citations or interactions among objects. This paper develops and estimates a simple consistent weighted ranking (CWR) scheme which, in the sports world, depends on four parameters (winning vs. losing and the relative importance of home vs. away games). In most ranking problems, there are not explicit criteria to evaluate the success of proposed rankings. NCAA college football has a special structure that enables the evaluation of each ranking scheme. Each season is essentially divided into two parts: the regular season and the post season bowl games. If a ranking scheme is accurate it should correctly predict a relatively large number of the bowl game outcomes. We use this structure to estimate the four parameters of our ranking function using "historical" data from the 1999-2003 seasons.

JEL Classification: C8 and L83
Keywords: NCAA football, ranking and sport tournament

Itay Fainmesser
The Eitan Berglas School of Economics
Tel Aviv University
Tel Aviv 69978
ISRAEL

Email: fainmass@post.tau.ac.il

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=163206

Chaim Fershtman
The Eitan Berglas School of Economics
Tel-Aviv University
Tel Aviv 69978
ISRAEL

Tel: (972 3) 640 7167
Fax:(972 3) 640 9908
Email: fersht@post.tau.ac.il

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=104930

Neil Gandal
Department of Public Policy
Tel Aviv University
Tel Aviv 69978
ISRAEL

Tel: (972 3) 640 6742
Fax:(972 3) 640 7382
Email: gandal@post.tau.ac.il

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=144974

Submitted 22 August 2005

1. Introduction

At the end of the regular season, the two top NCAA college football teams in the Bowl Championship Series (BCS) rankings play for the "so-called" national championship. Nevertheless, the 2003 college football season ended in a controversy and two national champions: LSU and USC. At the end of the regular season Oklahoma, LSU and USC all had a single loss. Although both the final regular season Associated Press poll of writers and ESPN/USA Today poll of football coaches ranked USC #1, the computer ratings were such that USC ended up #3 in the official BCS rankings. Although LSU beat Oklahoma in the "championship game," USC (which won its bowl game against #4 Michigan) was still ranked #1 in the final (post bowl) AP poll.[2]

Following the 2004 season, the BCS system again came under scrutiny; there were two complaints in 2004. The first involved Auburn, an undefeated team from a strong football conference, which was excluded from the BCS championship game. The only problem was that two other teams who play in strong conferences (USC and Oklahoma) were also undefeated. The other complaint involved California (Cal) which appeared to be on the verge of its first Rose Bowl appearance since 1959. The "Cal" controversy was due to the changes in the polls over the last week of the season. In the BCS ranking released following the week ending November 27, 2004, Cal was ranked ahead of Texas. There were only a few games the following Saturday, the last day of the regular season. Cal played and won while Texas did not play. Despite this, Cal fell and Texas gained in both the AP and USA Today/ ESPN polls and Cal lost its at-large invitation to a BCS bowl game.

The "disagreement" between the polls and the computer rankings following the 2003 college football season led to a modification of the BCS rankings that reduced the weight of the computer rankings. The "Cal" controversy following the 2004 season led, in part, to the AP poll dropping out of the BCS ranking system.

[2] By agreement, coaches who vote in the ESPN/USAToday poll are supposed to rank the winner of the BCS championship game as the #1 team. Hence LSU was ranked #1 in the final ESPN/USA Today poll.

Why is there more controversy in the ranking of NCAA college football teams than there is in the ranking of soccer teams in a European soccer league? In a setting in which each team plays against a small subset of the other teams and when teams potentially play a different number of games, ranking the whole group is nontrivial.[3] If we just add up the wins and losses, we only obtain a partial (and potentially distorted) measure. Some teams may play against strong teams while others may play against weak opponents. Clearly wins against high-quality teams cannot be counted the same as wins against weak opponents. Moreover such a measure will create an incentive problem; each team would prefer to play easy opponents. So there is clearly a need to put weights on each win (or loss) that will reflect the strength of the opponent.

Similar issues arise whenever one wants to establish ranking of academic journals, articles, patents, etc.[4] In all these cases the raw data for the complete ranking are bilateral citations or interactions between objects, individuals, or teams. Aggregating these interactions can be accomplished by the number of victories or a citation count, but typically it also requires some weighting function that captures the strength of the opponents or the importance of the citing articles or patents.

The weights in the ranking function can be given exogenously, for example when there is a known "journal impact factor" or a previous ranking of teams. But in such cases, the ranking function is not necessarily consistent. Ideally, the evaluation of each game or citation should be weighted by its importance, which is endogenously determined by the ranking procedure itself. A consistent ranking requires that the outcome of the ranking be identical to the weights that were used to form the ranking. This consistency requirement was first employed by Liebowitz and Palmer (1984) when they constructed their academic journal ranking.[5]

---

[3] There are 117 NCAA division I-A teams and each typically plays between 12-14 games, while in European soccer leagues, all teams play against each other.

[4] See for example Jaffe, and Trajtenberg (2002) for the use of patent citation data in examining the patern of knowledge spillovers and the dynamics of innovations. See Posner (2000) for a discussion of the role of judicial citations in the legal profession.

[5] See also Palacios-Huerta, I., and O. Volij, 2004 for an axiomatic approach for determining intellectual influence and in particular academic journal ranking. Their invariant ranking also satisfies the consistency requirement.

In the case of patents or journals articles, the outcomes are quite simple: either there is a citation or there is no citation. The problem is more complex in the case of sports rankings. The outcomes of a game include winning, losing, not playing, and in some cases, the possibility of a tie.[6] An analogy for wins and losses exists also for the case of academic papers. One could use data on rejections and the eventual acceptance of different papers in a similar fashion to wins and losses. A paper that was accepted by the AER without ever being rejected would be treated differently than a paper that was rejected by several other journals before it was accepted by the AER.[7]

In addition to the large set of possible outcomes, the location of the game may affect the outcome as well. This means that in addition to providing weights for the relative importance of wins vs. losses, weights must also be employed for the importance of "home games" vs. "away games".

This paper presents a simple consistent weighted ranking (CWR) scheme to rank agents or objects in such interactions and applies it to NCAA division 1-A college football.[8] In most ranking problems, there are not explicit criteria to evaluate the success of proposed rankings. NCAA college football has a special structure that enables the evaluation of each ranking scheme. Each season is essentially divided into two parts: the regular season and the post season bowl games. We use this structure twice; first we estimate the four parameters of our ranking function (winning vs. losing and the relative importance of home vs. away games) using "historical" data from the 1999-2003 seasons. The estimation procedure is quite straightforward. For each vector of parameters, the procedure uses the regular season outcomes to form a ranking among the teams for each season. If a ranking is accurate it should correctly predict a relatively large number of bowl game outcomes. Our procedure searched for the parameters that gave rise to the

[6] The score itself is part of the outcome as well.
[7] This is a hypothetical example since such data are not available.
[8] We are, of course, not the first to examine this issue. The wealth of information and rankings available at http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/index.shtml suggests that the rating of college football teams attracts a great deal of attention. Fair (2002) compares the relative predictive power of the BCS ranking schemes.

best overall score for the five seasons.[9] The estimated parameters suggest that a firm should be heavily penalized for a home loss and it indeed matters to whom one loses, that is, simply counting the number of losses does not lead to an accurate rating.

Finally using the estimated parameter, we employ the CWR in order to determine the ranking for the 2004 season. We then evaluate our ranking scheme by using it to predict the outcome of the 2004 post season (bowl) games. Our CWR ranking scheme predicted more bowl game outcomes correctly than the computer rankings used in the BCS rankings for 2004 period. While this is reassuring, this is only a one year comparison and clearly not statistically significant evidence regarding the quality of the various rankings. On the other hand the forecasting ability of our CWR scheme should improve as more seasons (data) are included in the estimation stage.

2. The BCS Controversies

Unlike college basketball (and other sports), there is no playoff system in college football. Hence, it was not always easy for the coaches' and writers' polls to agree on a national champion or the overall ranking. The BCS rating system which employs both computer rankings and polls was first implemented in 1998 to address this issue and try to achieve a consensus national champion.

As explained above, the 2003 college football season ended in controversy and two national champions: LSU and USC. The polls rated USC #1 at the end of the regular season, but only one of the seven computer formulas included in the 2003 BCS rankings had USC among the top two teams. While all three teams had 1 loss, the computer rankings indicated that Oklahoma and LSU had played a stronger schedule than USC.

The "disagreement" between the polls and the computer rankings inevitably led to comparisons of human and computer rankings as the following colorful quotation indicates.

---

[9] One can clearly think about alterative procedures; we discuss the issue in section 4.

> "To err is human. To err divinely takes a computer. Tarot cards, Ouija boards and Punxsutawney Phil could have done a better job than the Bowl Championship Series whizzes in picking teams for the national college football championship game… If the BCS computer formula had been used in the last presidential election, George Bush and Al Gore both would have lost and Rush Limbaugh would be in the White House on the strength of his radio ratings."[10]

The disagreement between the polls and the computer rankings also led to a modification of the BCS rankings following the 2003 college football season. The 2004 BCS rankings were based on the following three components, each with equal weights:[11]

- The ESPN/USA Today poll of coaches
- The Associated Press poll of writers
- Six computer rankings

Previously, the computer rankings made up approximately 50 percent of the overall BSC ratings. Hence there is a sense that the computer rankings were demoted. If the new system had been used during the 2003 season, LSU and USC would have played in the 2003 BCS championship game.

Following the 2004 season, the BCS system again came under scrutiny; the first complaint involved Auburn, which was excluded from the BCS national championship game, even though it was undefeated in 2004. Given that there were three undefeated teams, it is hard to fault the BCS here. No matter what system is used, only two of the three teams can play in the championship game. Given three undefeated, one team has to be excluded and that team will always feel slighted.

The other complaint involved California (Cal) which appeared to be on the verge of its first Rose Bowl appearance since 1959. Despite Cal's victory in its final game, Cal fell from 4th to 5th in the final BCS standings and lost its place to Texas, which climbed to

---

[10] The quotation comes from "No simple answer to BCS controversy," by Steve Wilstein, The Associated Press, available at the MSNBC website (http://www.msnbc.msn.com/id/3677199/).

[11] See http://www.bcsfootball.org/news.cfm?headline=40 for details. The NY Times withdrew from the BCS computer rankings after the 2003 season. The highest and lowest computer rankings for each team were discarded in the 2004 ranking BCS system.

number 4, despite being idle the final weekend. Texas thus obtained the BCS' only at-large berth and an appearance in the Rose Bowl, and Cal lost its place in a BCS bowl game.

The first question one might ask is given that there are 4 BCS bowl games (the championship game plus three additional BCS bowl games), how could the number 5 team in the country not be included? Common sense suggests that the BCS (the "C" stands for championship after all) would include the top 8 teams playing in the 4 BCS bowl games. Hence what follows is not to be taken as a criticism of Texas. Indeed if the BCS took the top eight teams, both Cal and Texas would have played in BCS games, perhaps against each other in the Rose Bowl.

But the BCS does not work that way. The conference champions of the six major conferences (SEC, Big 12, Big 10, Pac 10, ACC, and Big East) all qualify for BCS bowl games regardless of the final regular season rankings of those conference champions. In 2004, two of the conference champions were rated outside of the top 8. Hence Michigan, ranked number #13, and Pittsburgh ranked #21, appeared in BCS games. That would still appear to leave two at large spots. But Utah from a mid major conference was guaranteed a BCS bowl appearance by virtue of finishing in the top 6. (Utah finished number 6, so it would have appeared in a BCS bowl game if the top 8 teams had been chosen.) Hence, Texas and Cal were in the battle for the final BCS spot.

The controversy was due to the changes in the polls over the last week of the season. In the BCS ranking released following the week ending November 27, Cal was ranked ahead of Texas. There were only a few games the following weekend. Cal played December 4 against Southern Mississippi because an earlier scheduled game between the teams had been rained out by a hurricane. Cal beat Southern Mississippi on the road 26-16,[12] while Texas did not play. Despite Cal's victory and the fact that Texas did not play, Cal fell and Texas gained in both the AP and USA Today/ ESPN polls. The BCS computer ranking of the two teams were unchanged between the November 27 and

[12] Southern Mississippi finished the regular season 6-5 and later won its bowl game.

December 4 period. If there had been no changes in the polls, Cal would have played in the Rose bowl. The following table summarizes the changes that occurred in the polls and computer rankings between November 27 and December 4.

Games through	November 27	December 4	Actual Change (% change)
Polls			
Cal (AP)	1410	1399	-11 (-0.8%)
Texas (AP)	1325	1337	+12 (+0.9%)
Cal (ESPN/USA)	**1314**	**1286**	**-27 (-2.2%)**
Texas (ESPN/USA)	1266	1281	+15 (+1.2%)
BCS Computer Ranking: No change in California's and Texas' rankings			
Games: California 26 Southern Mississippi 16; Texas (idle)			

Although it is tough to explain Cal's slight drop and Texas' slight rise in the AP (sportswriters' polls), it is even harder to explain Cal's significant drop in the ESPN/USA Today (coaches' poll). Cal was ranked 7th by four coaches and 8th by two coaches in the poll taken after the December 4 game. This is despite the fact that no coach rated Cal lower than 6th in the poll following the November 27 games. The point is not the relatively low rankings themselves, rather the significant drop—after winning![13]

According to "Coaches Holding the Line on Keeping Polling Secret," Vittorio Tafur, *NY Times*, December 8th, Pac-10 conference commissioner Tom Hansen questioned the integrity of the coaches poll. The article notes that the teams in the Big 12 conference would split the $4.5 Million that Texas earned by virtue of reaching a BCS bowl game and that the payoff from the Holiday Bowl where Cal would appear was $2 million. Hence, it is not just the so-called national championship that it at stake. Huge financial windfalls typically ride on the BCS rankings since playing in a BCS bowl can result in millions of dollars for a school or conference,[14] while playing in a minor (non BCS) bowl

[13] Cal's ranking in the BCS computer polls did not change over this period; it was ranked #7 both before and after the final weekend of the regular season.
[14] According to the Outback bowl (http://www.outbackbowl.com/facts/collegegames.html), the 28 bowl games following the 2003 season distributed approximately $190 million to NCAA schools.

typically means much smaller payouts for the schools involved.[15] The article notes that when coaches vote on issues that affect the standings of their teams or teams in their conference, there is at the very least a perceived conflict of interest.

In part because of the "Cal" controversy following the 2004 season, the AP announced that it will no longer allow its poll to be used in the BCS rankings and ESPN withdrew from the coaches' poll. That leaves the coaches' poll (minus ESPN) and the computer rankings. To put it mildly, the BCS is in flux. Although the BCS is thinking of adding another poll,[16] the solution might be to give more importance to computer rankings. Despite the criticism of computer rankings, they are the only ones that can be transparent and based on measurable criteria, which is to say, impartial.

3. The CWR Ranking Methodology

We develop our ranking in three steps. We first consider a simple bilateral interaction like citations (cited articles or patent citations). This is relatively a simple case because either object i cites object j or it does not cite object j. We then consider a sports setting; in this case, there is a winner and a loser (or no game).[17] In the final stage we incorporate the possibility of two types of games; home games and away games.[18] This means that winning (or losing) a home game can have a different weight than winning (or losing) an away game.

3.1 Consistent Ranking

Consider a group $N \equiv \{1,....,n\}$ of agents (or objects), with the relation $a_{ij} \in \{0,1\}$ for every $i, j \in N$. For example, if N is a set of patents or articles, $a_{ij} = 1$ if patent or article j cites patent (or article) i and $a_{ij} = 0$ otherwise. Our dataset is hence uniquely defined by the

---

[15] In addition to the monetary payoffs a school receives for playing in an important bowl game, there are claims that donations to universities increase and the demand for attending a university increases in the success of the football team. Frank (2004) finds no statistical support for this claim.

[16] See "Notes: BCS could replace AP poll next month," available at http://www.usatoday.com/sports/college/football/bowls/2005-05-17-bcs-poll_x.htm?POE=SPOISVA.

[17] In hockey and soccer there is also the possibility of a tie that further complicates the analysis.

[18] The ranking of journals could have more structure if there were data on rejections of papers by different journals.

matrix $A=\lfloor a_{ij} \rfloor$. We interpret each $a_{ij}=1$ as a positive signal regarding object i. The objective is to define a function: $R: A \to R^n$ which generates a rating for every agent that summarize the information in A.

There are many possible ways to define the function R; the most trivial is the summation $r_i(A)=\sum_{j \neq i} a_{ij}$, $i=1,\ldots,n$, which is just a count; an example is the number of citations that each article receives. The advantage of such a ranking is its simplicity but it ignores much of the information embodied in A. Such a ranking may be appropriate when the "interactions" between the objects are not important; for example, when ranking bestsellers, a simple count of sales is probably appropriate. In other situations the identity or the "importance" of *j* should be taken into account when aggregating the a_{ij}. For example, in forming a ranking based on citations one may want to take into account the "importance" of the citing article.

One possible resolution is achieved by using an exogenous weighting vector, describing the agents' "importance." Examples include "Journal Impact Factors" or the use of polls (or previous rankings) in college football. Letting m_i be agent's *i* subjective significance, we can normalize the count in the following way:

$$r_i(A,m)=\sum_{j \neq i} m_j a_{ij} \quad , \quad i=1,\ldots,n$$

However, this ranking function is not "consistent". The rating used to determine each agent's influence (m_j) differs from the final rating (r_j) of the agents. This "inconsistency" can be fixed by requiring that the weight given to each a_{ij} is identical to the rating itself (see Liebowitz and Palmer [1984]), i.e.

$$z_i(A,z)=\sum_{j \neq i} a_{ij} z_j \ .$$

To guarantee uniqueness, we can employ a simple normalization requiring that $\Sigma z_i = 1$ and $\min_{i=1,\ldots,n} z_i = 0$. Specifically,

(1) $$z_i(A,z) = \frac{\sum_{j \neq i} a_{ij} z_j + g}{\sum_i \left(\sum_{j \neq i} a_{ij} z_j + g \right)}, \quad i = 1,\ldots,n, \text{ where } \min_{i=1,\ldots,n} z_i = 0,$$

where g is endogenously determined in order to enable a solution to the system (i.e., it is determined by the condition $\min_{i=1,\ldots,n} z_i = 0$). In order to solve system (1) we need to simultaneously determine the ratings of all agents (and g), since the ratings themselves are also the weights needed in the calculations.

3.2 Incorporating Wins and Losses

Our discussion up to this point considered the case when $a_{ij} \in \{0,1\}$. But in a sports match, the outcome can be win, lose, or do not play. Teams also might play more than one game against each other. To accommodate this we modify the ranking in the following way: For every $i, j \in N$, $a_{ij} \in Z^+$ indicates the number of times team i won against team j and $\overline{a}_{ij} \in Z^+$ indicates the number of times team i lost to team j, so the matrix $\overline{A} = \left[\overline{a}_{ij}\right]$ is added to the dataset (the matrix A is defined as above).[19] Returning to the analogy of ranking articles using both acceptance and rejection data, the $\overline{A}$ matrix would be the "rejection" matrix.

As before, our objective is to define a consistent ranking function $R : \langle A, \overline{A} \rangle \to R^n$. Allowing for different coefficients for wins and losses, equation (1) now becomes:

(2) $$z_i(A, \overline{A}, z) = \frac{\sum_{j \neq i} a_{ij} z_j - b \sum_{j \neq i} \overline{a}_{ij} (\gamma - z_j) + g}{\sum_i \left(\sum_{j \neq i} a_{ij} z_j - b \sum_{j \neq i} \overline{a}_{ij} (\gamma - z_j) + g \right)}, i = 1,\ldots,n, \quad \min_{i=1,\ldots,n} z_i = 0.$$

---

[19] Note that for every i,j $\overline{a}_{ij} = a_{ji}$, therefore there is no necessity in defining the new matrix $\overline{A}$. However, it will make the presentation of the system of equations clearer, especially when we introduce further extensions.

The variables b, γ are parameters that account for the importance of losses relative to wins. As b and γ increase, the rating gives higher weight to losses. The parameter γ has an additional interpretation; keeping $b \cdot \gamma$ constant, a large γ means that our ranking function primarily depends on the number of losses, while a small γ implies that the ranking is sensitive to whom one loses. To insure that winning increases a team's rating and losing decreases a team's rating, it must be the case that $b > 0,\ \gamma > \max_i z_i$.

3.3 Home Field Advantage

Sports experts claim that winning at "home" is easier than winning on the road. Since the location of the game is known, we can incorporate it in the ranking function by giving different weights to wins and losses at home and away games. We split each matrix $A(\overline{A})$, into home wins (losses) and away wins (losses). Thus, for every pair of teams $i, j \in N$, there are four relevant values $a_{ij}^{\text{home}}, a_{ij}^{away}, \overline{a}_{ij}^{\text{home}}, \overline{a}_{ij}^{away} \in Z^{+}$ which (respectively) describe the number of times team *i* won at home, won away, lost at home, and lost away, against team *j*. The four data matrices are: $A^{\text{home}}, A^{away}, \overline{A}^{\text{home}}, \overline{A}^{away}$. The ranking function is defined as follows:

(3)

$$z_i\left(A^{\text{home}}, A^{away}, \overline{A}^{\text{home}}, \overline{A}^{away}, z\right) =$$

$$= \frac{\left[\sum_{j \neq i} a_{ij}^{away} z_j + h^w \sum_{j \neq i} a_{ij}^{\text{home}} z_j\right] - b\left[\sum_{j \neq i} \overline{a}_{ij}^{away}\left(\gamma - z_j\right) + h^l \sum_{j \neq i} \overline{a}_{ij}^{\text{home}}\left(\gamma - z_j\right)\right] + g}{\sum_i \left(\left[\sum_{j \neq i} a_{ij}^{away} z_j + h^w \sum_{j \neq i} a_{ij}^{\text{home}} z_j\right] - b\left[\sum_{j \neq i} \overline{a}_{ij}^{away}\left(\gamma - z_j\right) + h^l \sum_{j \neq i} \overline{a}_{ij}^{\text{home}}\left(\gamma - z_j\right)\right] + g\right)}$$

Again, $\min_{i=1,..,n} z_i = 0$.

Road wins and road losses are normalized to one. Hence the parameters h^w and h^l account for the weight of home wins (losses) relative to away wins (losses) in calculating the ratings.

4. Estimation and Evaluation of Ranking Parameters

Equation (3) is our ranking function, but it requires an input of four exogenous parameters: b, γ, h^w , and h^l . Determining the values of these parameters might be considered a task for football analysts. We clearly do not claim to possess such expertise. Instead, we propose to estimate these parameters using data from previous seasons.

The NCAA college football season is set up in a unique way that facilitates the evaluation of different ranking schemes. There are essentially two rounds in the college football season. In the first round, there are regular season games; in the second round, there are the so-called bowl games. Teams that play well during the regular season are invited to bowl games.

This setting provides us with a natural experiment to test the different ranking schemes. The regular season ranking determines the relative strength of the team. The performance of each ranking can be evaluated by its implied prediction of the bowl game outcomes. If a ranking is reasonably good, then in a bowl game involving the #3 and #9 teams, the probability that the team ranked #3 wins the game should be more than 50%. We can thus use the results of the bowl games to evaluate the quality of the pre-bowl rankings.

The bowl games are only for the better teams. Since we use these games in determining the parameters, our ranking may not be that accurate for the teams below the median. (Approximately 50% of the teams participate in bowl games.) While this may be a positive feature, it means that more caution should be used when comparing the rankings of the lower ranked teams.

We use the 1999-2003 seasons to estimate the parameters: b, γ, h^w and h^l.[20] For a given set of parameters, there is, for every year, a unique pre-bowl rating. The second step is to examine the bowl games and determine which parameters provide the best prediction. There are clearly different ways to evaluate the performance of each rating system; we adopt a simple rule that selects the parameters that predict the highest number of bowl game results correctly over the five year period.[21]

For every set of parameters we assign a grade G(b, γ, h^w, h^l) which equals the number of bowl games (during the 1999-2003 period) predicted correctly by the ranking derived from these parameters. A correct prediction means that the winner of the bowl game is the higher ranked team at the end of the regular season. Fortunately bowl games are played at neutral sites (i.e., no home field advantage for either team) so the prediction of the outcome of the bowl games depends only on the teams' relative ranking. Following the 1999 season there were 24 bowl games, following the 2000-2001 seasons there were 25 bowl games each year, while following the 2002-2003 seasons there were 28 bowl games each year. Thus the maximum overall score for the 1999-2003 period is 130, the number of bowl games during that period. We then sum up the number of correct predictions for the five years of bowl games associated with each set of parameter estimates. This gives us a grade, G(b, γ, h^w, h^l), for every set of parameters.

We first chose relatively broad intervals for the parameters in order to find areas which provided the best grade. The values chosen for the initial grid (see Table 1 below) were not chosen randomly: b which accounts for the importance of losses relative to wins was allowed to vary between 0.1 and 2.8. This means that the importance of losses relative to wins could vary between 10% and 280%. γ was allowed to vary between from 0.02 to 0.32. A γ of 0.32 is roughly 15 times the rating of the most highly ranked team; hence the range for γ is also very large. h^w and h^l were chosen to allow a large range as well.

[20] Some of the bowl games of the 2003 season, for example, take place in early January 2004. For ease of presentation we refer to them as games of the 2003 season.
[21] We discuss alternatives below.

	b	γ	h^w	h^l
Lower bound	0.1	0.02	0.1	0.1
Upper bound	2.8	0.32	2.8	2.8
Broad Grid intervals	0.3	0.05	0.3	0.3
Narrow Grid intervals	0.1	0.01	0.1	0.1

Table 1: Initial Grid and Intervals

Using the results from the initial grid, we changed and narrowed parameter range and increased the resolution around two distinct areas that yielded high grades.[22] The best predictions were given by two sets of parameters in two areas of the grid; these two distinct areas yielded 82 correct predictions over the five year period (out of a possible 130), or 63%. The following two sets of parameters are at the center of the two regions with the highest scores:

Parameters	b	γ	h^w	h^l
Set 1	1.1	0.03	0.9	1.8
Set 2	1.9	0.03	2.6	1.6

Table 2: Optimal Parameter Sets

In the first set of parameter estimates, h^w<1 while h^l>1. The large difference between h^w and h^l suggests that a team is heavily penalized for a home loss, relative to a road loss (which is normalized to one) and that a home win is rewarded only slightly less than a road win (which is normalized to one). For this set of parameters, losing at home is a key factor in assessing a team's rating. When b is close to 1, this suggests that wins and losses affect the ratings symmetrically. b=1.1 suggests that ratings are slightly more sensitive to losses than wins.

In order to interpret γ, we need to know that the highest rating in 2004 was approximately 0.02. This means that other things being equal, the "loss penalty" from losing to a very

[22] The search algorithm was written in Matlab. The algorithm and the complete set of results for the whole broad and narrow grids are available upon request. See the appendix for the shape of the objective function.

highly rated team is γ - .02= .01, which is 1/3 the "loss penalty" of losing to a team with a very low rating (γ - 0 = .03). Hence, the relatively low γ suggests that it indeed matters to whom one loses. (A relatively high γ implies that the ranking is more sensitive to the number of losses, rather than to whom one loses.)

In the second set of parameters, b and h^w are both higher, while γ and h^l are essentially unchanged relative to the first set of parameters. Hence in both cases, the estimated parameters suggest that a firm should be heavily penalized for a home loss (h^l is relatively large) and it indeed matters to whom one loses (γ is relatively small).

The two different sets of parameters give similar results because of the substitutability among b and h^w. For example, as b rises from 1.1 to 1.9, more weight is given to losses relative to wins. This effect is offset in large part by a higher value of h^w (0.9 in the first set of parameters and 2.6 in the second set of parameters), which increases the importance of the home wins.[23]

Alternative Estimation Methods: There are several possible ways to use the regular season ratings to forecast the bowl games results. We used the most straightforward method; the estimated parameters predicted the highest number of bowl game outcomes correctly. An alternative method is to use the rating of two teams (not the ranking) to predict the probability that team a beats team b in the bowl game. For example, if z_i, $i \in \{a, b\}$ is the rating of team i, then $\Pr\{a \text{ beats } b \mid z_a, z_b\} = \frac{z_a}{z_a + z_b}$. In order to evaluate the quality of a prediction of a given rating for the bowl games, one could then use a least squares method. The objective function to be minimized would then be $\sum_{b \in N} \sum_{a \in \{a \mid a \in N, a \text{ beat } b\}} \left(1 - \frac{z_a}{z_a + z_b}\right)^2$.

[23] In the appendix, we choose two relatively high and two relatively low values for each parameter in order to provide a sense as to the shape of the objective function.

This method uses more data than the method we chose since it exploits the whole (cardinal) rating, while our method relies solely on the (ordinal) ranking. Nevertheless, the method we have chosen is more intuitive and enables us to compare our results with the other ranking systems. (For other ranking systems, we only have data on rankings and not on the ratings.) Furthermore, as we noted, the BCS uses the average of the computer rankings, not the ratings. Therefore, the goal of the different computer ratings is likely to create a reasonable (ordinal) ranking. It might not be fair to compare them on a cardinal basis even if we had such data.

5. Evaluating the Performance of the Different Ranking Schemes for the 2004 Season

Finally, we now compare our ranking methodology with the rankings of the experts. We use the 2004 season – which was not used in estimating the parameters of the ranking – and perform a simple test. Using the parameters that we estimated in section 4, we rank the teams for the 2004 regular season. We then calculate the number of bowl games whose outcomes were correctly predicted following the 2004 regular season; we compare our result with the number of correct predictions from the six computer ranking schemes employed in the BCS ranking.

The six computer rankings included in the 2004 BCS rankings are:[24]

- AH- Anderson & Hester ratings (http://www.andersonsports.com/football/ACF_SOS.html),
- RB - Richard Billingsley ratings (http://www.cfrc.com/),
- CM - Colley Matrix ratings (http://www.colleyrankings.com/matrate.pdf),
- KM - Kenneth Massey ratings (http://www.mratings.com/rate/cf-m.htm),
- JS – Jeff Saragin ratings, (http://www.usatoday.com/sports/sagarin.htm),
- PW - Peter Wolfe ratings (http://www.bol.ucla.edu/~prwolfe/cfootball/ratings.htm).

Several of these computer rankings seem consistent, although it is difficult to make definitive conclusions about consistency in all instances; only a couple of the algorithms

[24] There are many other computer rankings in addition to the six used by the BCS. Massey, for example, includes 97 rankings on his comparison page. See, for example, the ratings comparison page at the end of the regular season in 2003, available at http://www.masseyratings.com/cf/compare2003-15.htm. Homepages for these rankings can be found at http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/index.html.

are discussed in detail. To the best of our knowledge, none of these systems use the results of the bowl games in previous years to determine the input parameters for the rankings. As Table 3 indicates our methodology predicted 15 or 16 out of 28 bowl games correctly in 2004, while the six computer programs used by the BCS predicted between 10-14 games correctly.

Ranking	CWR 1	CWR 2	AH	CM	KM	RB	PW	JS
# of correct predictions	15	16	12	13	11	14	10	14
% of correct predictions	0.54	0.57	0.43	0.46	0.39	0.50	0.36	0.50

Table 3: Bowl Games Predicted Correctly for the 2004 Season[25]

We should add a word of caution here: while these results are interesting, they do not necessarily suggest any significant difference between our ranking schemes and those of the computer ranking schemes used by the BCS since the comparison is only for a single year.

In Table 4, we report the number of correct predictions for the ranking schemes for the 1999-2003 seasons as well. This comparison is, of course, somewhat unfair, because our optimization methodology chose the parameters that led to the highest number of correctly predicted bowl games during the 1999-2003 period. Nevertheless, it suggests that there may be benefits from using historical data in developing ranking schemes.

[25] CWR 1 refers to the first set of parameters discussed in section 5, while CWR 2 refers to the second set of parameters.

Ranking	CWR 1	CWR 2	AH	CM	KM	RB	PW	JS
1999	15	17	14	12	14	NA[26]	NA	NA
2000	17	16	15	13	12	16	NA	NA
2001	15	16	14	14	15	12	NA	NA
2002	17	17	14	15	13	13	14	13
2003	18	16	15	15	19	21	17	19
Total 1999-2003	**82**	**82**	**72**	**71**	**73**	**NA**	**NA**	**NA**

Table 4: Bowl Games Predicted Correctly for the 1999-2003 Seasons

6. Concluding Remarks: The BCS Controversies Revisited

We now briefly discuss how the controversies would have had played out had our ranking scheme been used with the first set of estimated parameters in Table 2. In the case of the Cal controversy, our ranking had Cal ranked #9 before the final week of the season, just slightly below #7 Iowa and #8 LSU. Following its victory over Southern Mississippi (and taking account of the results of the other games that weekend), Cal climbed to #7 in the final regular season rankings. The computers also had Cal ranked #7 immediately before the final week of the season and #7 at the end of the regular season.[27] Hence the computers seem to be consistent; a victory led to a higher ranking in our case and an unchanged ranking in the case of the average of the BCS computer rankings. The polls on the other hand demoted Cal following its victory over Southern Mississippi.

In the 2003 controversy – three top teams with one loss – we had Oklahoma (#2) and USC (#3) ranked above LSU using the first set of parameters in Table 2. But our top team in 2003 was Miami of Ohio. Is that absurd? Miami lost only once in 2003, on opening day to Iowa which finished the year 13th in the final BCS rankings. The average of the BCS computer rankings had Miami of Ohio ranked #6 at the end of the regular season. Additionally, Miami defeated Louisville convincingly in the 2003 GMAC bowl game. Finally, Miami's 2003 quarterback Ben Roethlisberger went on to lead the Pittsburgh Steelers to the best regular season record (15-1) in the NFL in 2004. In 2003

[26] NA= Data Not Available.

[27] The second set of parameters gives qualitatively similar changes. With its win over Southern Mississippi, Cal climbed from #10 to #9 with this set of parameters.

without Roethlisberger the Steelers were 6-10. Perhaps it's not so absurd that Miami of Ohio was our top team in 2003! Regarding the 2004 "three undefeated teams" controversy, we had Oklahoma and Auburn ranked #1 and #2 respectively, not unreasonable given Auburn's tough schedule.[28]

References

Fair R., and J. Oster, 2002 "Comparing the Predictive Information Content of College Football Ratings," mimeo, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=335801.

Frank, R., 2004 "Challenging the Myth: A Review of the Links Among College Athletic Success, Student Quality, and Donations," Knight Foundation, executive summary available at http://www.knightfdn.org/default.asp?story=athletics/reports/2004_frankreport/summary.html.

Jaffe, A., and M. Tratjenberg, 2002, Patents, Citations, and Innovations, MIT Press, Cambridge, MA.

Liebowitz, S. and J. Palmer (1984), "Assessing the relative impacts of economic journals" *Journal of Economic Literature*, 22, 77-88.

Palacios-Huerta, I., and O. Volij, 2004 "The Measure of Intellectual Influence," *Econometrica*, 72: 963-977.

Posner, R. A. (2000) An economic analysis of the use of citation in the law" *American Law and Economic Review* 2(2), 381-406.

[28] With the second set of parameters, Oklahoma was #1, Miami of Ohio #2, and USC #3 in 2003, and Oklahoma and Auburn were (again) #1 and #2 respectively in 2004.

Appendix: Shape of the Objective Function

In this appendix, we choose two relatively high and two relatively low values for each parameter in order to provide a sense as to the shape of the objective function. The "low" parameters employed in the table below are b=0.7, γ=0.03, h^w=0.7, h^l=0.7, while the high parameters are b =2.2, γ=0.27, h^w=2.8, h^l=2.8. Grade refers to the number of correct predictions for the 1999-2003 seasons.

Parameters				**Grade**
b	**γ**	**h^w**	**h^l**	
High	High	High	High	49
High	High	High	Low	58
High	High	Low	High	52
High [29]	Low	High	High	81
Low	High	High	High	49
High	High	Low	Low	59
High	Low	High	Low	73
High	Low	Low	High	77
Low	High	High	Low	70
Low	High	Low	High	49
Low	Low	High	High	77
High	Low	Low	Low	71
Low	High	Low	Low	69
Low	Low	High	Low	73
Low	Low	Low	High	77
Low	Low	Low	Low	72

[29] This set of parameters roughly corresponds to the second set of parameters in Table 2.