Journal of Quantitative Analysis in Sports

Volume 3, Issue 2

2007

Article 2

Dimension Reduction for Hybrid Paired Comparison Models

David H. Annis*

*Wachovia Treasury, Dave@sportsquant.com

Copyright ©2007 The Berkeley Electronic Press. All rights reserved.

Dimension Reduction for Hybrid Paired Comparison Models

David H. Annis

Abstract

Rating – and subsequently ranking – college football teams requires making sense of sometimes conflicting pair-wise comparisons. Classical statistical techniques fall into one of two classes: win/loss models, which focus on binary outcomes, and point-scoring models, which consider the distribution of component scores. Annis and Craig (2005) illustrate deficiencies of both, and propose a hybrid method that considers both sources of data. Their method, while providing satisfactory results in many circumstances, can be difficult to implement numerically. This paper presents a refinement of their hybrid rating algorithm which preserves their original intent but simplifies greatly its implementation. Like its predecessor, the new model enjoys robustness to model misspecification, while offering substantial simplification.

KEYWORDS: paired comparisons, college football, hybrid model

1 INTRODUCTION

Annis and Craig [1] classify rating algorithms broadly into one of two distinct classes: win/loss models and point-scoring models. Win/loss models consider 0/1 observations representing losses and wins, respectively, but ignore any measure of degree. By reducing a game to a single binary outcome, they are susceptible to degenerate solutions (i.e., unbeaten teams will always be rated infinitely superior to others irrespective of schedule faced). Conversely, point-scoring models avoid degenerate solutions in all but the most pathological cases. However, under their canonical specification they suffer from a sufficiency problem¹ which causes them to devalue the importance of winning the game.

Their solution to these problems was a hybrid ranking procedure which models wins and losses as well as points scored. Owing to its generality², the Annis-Craig method is robust to model misspecification and performs well for a wide range of data generating mechanisms in which other approaches fail. However, this robustness and generality come at a computational price. In some circumstances – particularly when the sample size is small, as is the case in college football – their ranking algorithm can be difficult to implement due to the requirement that the win/loss and point scoring parameters be shared. This requirement is necessary to achieve a non-circular rank ordering and to keep the already large number of parameters (relative to the sample size) as small as possible.

Geometrically, the flexibility of the link functions required to map a common set of parameters to both scores and win probabilities can result in a flat (log-quasi-)likelihood surface. This occurs because the link parameters affect the scale of the regression parameters. Consequently, different combinations of links and regression coefficients can produce qualitatively similar fits despite drastically different parameter values. In these cases, iterative numerical routines can encounter difficulty optimizing the full formulation. Annis and Craig [1] note this in section 4.2. This, perhaps, explains the relatively infrequent use of embedded parametric links in applications of generalized linear models.

This work introduces an effective reduced-dimensional model that requires fewer link parameters (one vs. three) and half the number of team strength parameters (one overall strength parameter in lieu of separate offensive and defensive ones) than the original. The new, reduced-order model preserves the

¹For a given schedule, when a generalized linear model for points scored is fit using a canonical link, sufficient statistics for offensive and defensive abilities are total points scored and total points allowed.

²The Annis-Craig model is a generalization of many common win/loss and point-scoring models. For particular parameter values, their model reduces to the Bradley-Terry [4] and Thurstone-Mosteller [5], [7] models for binary data and Normal, Poisson and Gamma regression models for point-scoring data.

Annis-Craig spirit of differentiating between convincing wins and close ones. Simulation studies show it to be superior to win/loss and point-scoring models, while lagging, slightly, the Annis-Craig model.

The remainder of the paper is organized as follows. Section 2 introduces a reduced-dimensional hybrid model and discusses parameter estimation. Section 3 compares the performance of the new method to existing rating algorithms. Computational considerations, specifically reduction in run-times, are discussed in section 4, and some concluding remarks are made in section 5.

2 REDUCED DIMENSIONAL HYBRID MODEL

The primary advantage of the Annis-Craig algorithm is that it explicitly considers wins and losses, yet acknowledges that not all wins are equally impressive. This paper proposes a method which accomplishes that objective by dividing a unit score between the winning and losing teams. These quasi-binary outcomes take values along the entire interval $(0, 1)$ rather than only at the end-points, as would be the case for a pure win/loss model.

Of particular interest, therefore, is determining an appropriate, objective manner by which to assign credit. Intuitively, a team posting a convincing win should earn most of the unit, leaving next-to-nothing for the loser. On the other hand, both teams involved in a close game should earn roughly one-half unit. In other words, the credit earned should increase with the impressiveness of the performance.

2.1 MAPPING FUNCTION

An appropriate mapping function takes a score-pair into a fraction, z , of the unit interval, such that the winning team is awarded $\max(z, 1 - z)$ while the loser is awarded $\min(z, 1 - z)$. Because there are many possibilities, such a function is difficult to specify. Indeed, many intuitive formulations have serious drawbacks. For example, the proportion of points scored might work well for high-scoring sports such as basketball. However using a proportion suggests that a 3-0 (100%) win is more convincing than a 52-3 (94.5%) one. Another potential mapping function is (scaled) point differential, which implies that a 7-3 defensive struggle and a 42-38 shootout are equivalent. In light of these difficulties, rather than specifying the mapping function *a priori*, it will be estimated nonparametrically from the data subject to the following constraints:

1. The winning team must earn at least 0.5. In other words, winning matters. The winning team will always receive more credit than the losing team.

2. The mapping is monotonically non-decreasing in a team's points scored and monotonically non-increasing in points allowed.

Requirement 2 states that for a fixed number of points allowed, a team will not earn less credit by scoring more points. Conversely, for a fixed point total, a team can not earn more credit by allowing its opponents to score more. These monotonicity requirements suggest applying isotonic regression (see, e.g., [3]) techniques. However, classical isotonic regression is inappropriate for two reasons. First, there is no easy way to preclude values of $z \notin [0, 1]$. Second, the procedure assumes constant variance, which is questionable, since one touchdown makes a dramatic difference in a close game, but is relatively meaningless in a blowout. Both problems can be remedied by generalizing Bacchetti's [2] additive isotonic model, in which a monotonic function of z modeled as the sum of isotonic functions of the scores.

Ideally, one would observe team scores, s_1 and s_2 , and the earned credit, z , from each game. However, because z is latent, it is only possible to observe $I_{[z>1/2]}$, where $I_{[A]}$ denotes an indicator variable equal to 1 if event A occurs and 0 otherwise. Unfortunately, the estimated mapping function taking $\{s_1, s_2\}$ to $I_{[z>1/2]}$ is degenerate. To see this, consider a logit³ link for the response and the isotonic functions $m_1(s) = cs$ and $m_2(s) = -cs$ for $c \gg 1$.

$$\hat{z} = \frac{\exp\{m_1(s_1) + m_2(s_2)\}}{1 + \exp\{m_1(s_1) + m_2(s_2)\}} ; \quad \lim_{c \rightarrow \infty} \hat{z} = \begin{cases} 0, & s_1 < s_2; \\ 1, & s_1 > s_2. \end{cases}$$

As $c \rightarrow \infty$, all credit is given to the winning team and none to the losing team, matching the observed indicator functions exactly, but ignoring any measure of degree. This is demonstrated in Figure 1, which shows that the particular scores have no influence on the credit share, z .

An alternative approach is based on expected values of the particular quantities. The Annis-Craig model provides estimates of $E(s_i|\delta_{i,j})$, $E(s_j|\delta_{i,j})$ and $\pi_{i,j}$, which can be used as a proxy for z . Here, s_i and s_j are the scores for teams i and j , respectively; $\delta_{i,j}$ is an indicator variable equal to 1 if i defeats j and 0 otherwise; and $\pi_{i,j}$ is the probability that i defeats j . Estimated unconditional scores are computed, trivially, as

$$\begin{aligned} E(s_i) &= \pi_{i,j} \times E(s_i|\delta_{i,j} = 1) + (1 - \pi_{i,j}) \times E(s_i|\delta_{i,j} = 0) \\ \text{and } E(s_j) &= \pi_{i,j} \times E(s_j|\delta_{i,j} = 1) + (1 - \pi_{i,j}) \times E(s_j|\delta_{i,j} = 0). \end{aligned}$$

The isotonic map is then fit on the expected triple $\{E(s_i), E(s_j), \pi_{i,j}\}$. There are two advantages of using expectations in place of the observed values to model z . As previously noted, using observed scores and an indicator for z fails to capture any measure of decisiveness associated with a win. More

³The phenomenon persists for all monotone link functions.

Figure 1: The isotonic mapping of observed data is binary.

subtly, by using expected scores, it is possible to generate a much larger (in fact, infinite) “data set” on which to fit the model for z because any number of hypothetical match-ups between teams of user-defined abilities can be considered, not merely games between particular teams which have occurred.

The winning team’s share is modeled as an additive isotonic function of the winning and losing scores. A transformed variable, $z^* = (2z - 1)$, is taken as the response. Here z^* represents the proportion of the half-interval $(0.5, 1)$ that is awarded to the winning team in excess of the required minimum of 0.5. Because it is possible for an additive isotonic function of the untransformed response to award $z < 0.5$ to the winning team, this transformation is necessary to ensure requirement 1 is satisfied. Therefore $z^* \in (0, 1)$ for $z \in (0.5, 1)$. Further, it is assumed that the variance, $V(z^*)$, is proportional to $(1 - z^*)$, reflecting that the credit division in tight games ($z^* \approx 0$) is more uncertain than in blowouts ($z^* \approx 1$). An iterative, back-fitting approach similar to Bacchetti’s is used to model $\text{logit}(z^*) = m_w(s_w) + m_\ell(s_\ell)$, where $m_w(\cdot)$ and $m_\ell(\cdot)$ are monotonically non-decreasing and non-increasing functions of the winning and losing scores, respectively.

Additive isotonic mappings were estimated separately for the 2004, 2005 and 2006 seasons. The fits suggest no appreciable differences, indicating, as expected, that the scale on which scores are measured does not change appreciably year-to-year. Figure 2 gives a three-dimensional response surface for the

isotonic mapping based on 2005 data, which is representative of other years. The ridge along the main diagonal is desirable and occurs because of the discontinuity between winning and losing. However, in contrast to Figure 1, this discontinuity is much less pronounced and therefore allows sufficient gradation of results. The shape of the mapping in Figure 2 is appealing. Its slope is most

Figure 2: Response Surface for Additive Isotonic Mapping.

steep near the 45-degree line, indicating that a few points are very meaningful in those games than in lopsided contests. Furthermore, as the outcome deviates from the 45-degree line, the additional reward (or penalty) for a single point decreases. This suggests that while a 50-3 win is more convincing than a 49-3 win, it is only marginally so.

Earlier, it was argued that a 7-3 win was more convincing than a 42-38 win even though both were by a margin of four points. Figure 2 confirms the phenomenon in the data and validates this intuition. Using the estimated isotonic mapping, winning and losing teams in a 7-3 game receive 0.754 and 0.246 for their performances, respectively.

$$\begin{aligned}\logit(z^*) &= m_w(7) + m_\ell(3) = -4.512 + 4.547 = 0.035 \\ z^* &= \frac{e^{0.035}}{1 + e^{0.035}} = 0.509 \\ z &= \frac{z^* + 1}{2} = 0.754.\end{aligned}$$

Figure 3 illustrates the look-up of the isotonic function values. In this example, the relatively low value of $m_w(7) = -4.512$ indicates that, compared to other winning scores, 7 is quite small. On the other hand, most winning teams allow more than 3 points, and consequently $m_\ell(3) = 4.547$ is much larger than other values of the function. Similar calculations show the winning and losing teams

Figure 3: Isotonic functions of winning and losing scores determine game credit.

in a 42-38 game receive 0.512 and 0.488, respectively, due to the closeness of the outcome. Likewise, teams involved in a 31-27 game receive 0.601 and 0.399, indicating a more decisive outcome than 42-38, but a less decisive one than 7-3. This reinforces the intuition that point differential alone does not sufficiently capture the extent to which one team outplays another.

Figure 4 gives a contour plot for the estimated isotonic map from the perspective of team 1 in a game against team 2. Note the contours are not parallel to the 45-degree line. This “fanning out” of the contours indicates that as the losing score increases, the winning team’s point differential must increase for it to earn the same credit. Other interesting features are the flat areas in the lower left (where both scores are less than 4) and upper right (where both scores exceed 40 points). They illustrate that neither extremely low scoring games (in which the winning score is a field-goal or safety) nor extremely high-scoring games (in which the losing team scores over 40 points) are decisive. In other words, a winning team incapable of scoring more than 3 points or holding its opponent to under 40 has not distinguished itself.

Figure 4: Point differential is insufficient to describe the quality of a win.

2.2 PARAMETER ESTIMATION

This method assumes there is negligible change in the estimated mapping function from year-to-year, an assumption which is consistent with results from the 2004 through 2006 seasons. Therefore, in practice, it is only necessary to estimate the mapping function once to initially calibrate the model. It would be necessary to estimate different isotonic mapping functions for different sports, and even different levels of the same sport, e.g., college vs. professional football. However, examination of multiple seasons of college football data supports the intuition that the game remains relatively consistent year-to-year and therefore the mapping function need not be updated frequently. (Were this not the case, this proposed model would be of little use, as it would be necessary to fit the more numerically intensive model first before fitting the simpler one, thus negating the latter's computational advantage.) In addition, estimating the isotonic relationships requires only estimates of the link parameters from the Annis-Craig model. Individual team strengths are not required, since expected value triples can be calculated for any values of team strength in any number of hypothetical games. Conceivably, the calibration can be sped up further by considering only a carefully chosen subset of the original data set adequate for estimating the link parameters.

Given the mapping, each game can be reduced to a single quasi-binary response, $z \in (0, 1)$. Team ratings, on which rankings are based, are determined

by modeling $E(z)$ as functions of the teams involved and venue. Following Annis and Craig [1], a quasi-likelihood is assumed for the earned game credit. The first two moments are

$$E(z_{i,j}) = H(\alpha_i - \alpha_j \pm \psi) = \zeta_{i,j} \quad \text{and} \quad V(z_{i,j}) = \phi \zeta_{i,j}(1 - \zeta_{i,j}),$$

where

$$H(\eta|\kappa) = \int_0^\eta \frac{e^{t\kappa}(1 + e^t)^{-2\kappa}}{Be(\kappa, \kappa)} dt$$

is a symmetric version of Prentice's [6] binary link function, ϕ is an overdispersion parameter and ψ is an optional home-field advantage parameter. The restricted version of Prentice's link is required to ensure the requisite symmetry for a zero-sum game. That is, if team 1 receives z from a game against team 2, this is equivalent to stating that team 2 earned $(1 - z)$ from a game against team 1. Asymmetric links do not satisfy this property and are therefore inappropriate for analyzing paired comparisons. Concurrent estimation of team strength parameters, α , and the link parameter, κ , is straightforward and makes use of the gradient-based method laid out by Annis and Craig⁴.

3 SIMULATION STUDIES

Three simulation studies were conducted to assess the performance of the proposed ranking procedure. Each consisted of 100 randomly generated seasons. In the first, wins and losses were generated according to a Bradley-Terry formulation. In the remaining two, independent point totals were simulated from Normal and overdispersed Poisson distributions respectively⁵. In all cases, parameters were estimated under the appropriate model using data from the 2006 season, and ranks were derived from estimated team strengths such that the strongest team is ranked #1, the next strongest is ranked #2, etc. The isotonic mapping function was estimated from the 2005 season to examine robustness.

For each series of simulations, the proposed method is compared to five others based on three criteria. The alternative models are the Bradley-Terry and Thurstone-Mosteller models for win/loss data, the Normal and Poisson models for point-scoring data, and the Annis-Craig hybrid model. Model performance is judged by the percentage of times the correct #1 team is identified, the percentage of times the top two teams are correctly identified (irrespective of order) and the Spearman rank correlation between the true and estimated rank ordering of all teams.

⁴Annis and Craig use m in place of κ , however because m is used to denote the monotonic functions used in the mapping, κ is used here to avoid confusion.

⁵Normal variates were truncated at zero, and, when necessary, ties were broken by generating Bernoulli random variables with probability one-half.

To avoid potential confusion, in the following discussion “truth” refers to the *known parameter values* used to conduct the simulations. Although it would be preferable to consider “true” values based on observed games, it is impossible since team strengths can never be known, only estimated.

It is worth mentioning that these simulations – particularly the Bradley-Terry – are unfavorable for the proposed method. This is due to the reduced-dimension procedure’s reliance on a mapping function which is estimated from observed game data. Because these simulations are approximations to game outcomes, the isotonic mapping will perform worse in these situations than when applied to real data. Therefore, the simulations represent a conservative assessment of the procedure.

3.1 BRADLEY-TERRY SIMULATIONS

In the first study, wins and losses were generated using a Bradley-Terry specification. Subsequently, losing scores and margins of victory were sampled from their empirical distributions. Sampling in this manner avoids non-negative scores and ties. By construction, the only information relevant for differentiating teams is contained in the wins and losses; scores are meaningless.

As expected, the win/loss models outperform the point-scoring models, with both the complete and reduced hybrid models in between. However, because the reduced model was fit using an inappropriate mapping function, its performance here understates its true potential. Recall, the isotonic map is estimated using 2005 game data (for which scores are not meaningless). Clearly, these simulations are an extreme case and do not mimic college football accurately. Were this situation representative, the proposed technique would benefit substantially from re-estimating the mapping function. When this is done, the map resembles Figure 1, indicating (correctly) that the scores themselves are inconsequential and that credit awarded should be a step function differentiating between a win and a loss. When the mapping is re-fit, results differ insignificantly from those based on either win/loss model.

Despite this systematic disadvantage, the reduced hybrid formulation based on the 2005 isotonic map still outperforms point-scoring models. Thus, not only is the proposed method robust to model misspecification, it is, to a degree, robust to incorrect choices of the isotonic mapping function. The new procedure identified the #1 team and top two teams more frequently (13%) than either point scoring method (12% and 11% for the Normal and Poisson, respectively). As expected, the Bradley Terry and Thurstone-Mosteller models (39% each) were better in this case, as was the complete Annis-Craig model (39%). Similar results are noted for the Spearman correlations between the true and estimated rank-orders. Comparisons based on the percentage of times the top two teams were identified are inconclusive, but show the same

95% CI for difference from	Correct #1		Correct Top 2		Rank Correlation	
	Lower	Upper	Lower	Upper	Lower	Upper
Bradley-Terry	-0.2611	-0.2589	-0.2911	-0.2889	-0.0410	-0.0388
Thurstone-Mosteller	-0.2611	-0.2589	-0.2711	-0.2689	-0.0400	-0.0379
Normal	0.0092	0.0108	-0.0108	-0.0092	0.0480	0.0496
Poisson	0.0192	0.0208	0.0192	0.0208	0.0468	0.0484
Annis-Craig	-0.2611	-0.2589	-0.1611	-0.1589	-0.0407	-0.0384

Table 1: The reduced-dimension hybrid model outperforms point-scoring models when wins follow a Bradley-Terry specification.

general trend. Simultaneous 95% Bonferroni confidence intervals for the *difference* between the low-order hybrid model and all others for each metric are given in Table 1.

3.2 NORMAL SIMULATIONS

The second simulation generated independent, Normally distributed point totals. The reduced-order hybrid model identified the correct #1 team more often (47%) than all competing methods (35% for the Bradley-Terry, 40% for the Thurstone-Mosteller, 41% for the Normal, 25% for the Poisson and 45% for the Annis-Craig models). It is interesting that the reduced-dimension model outperforms the true model and the more general Annis-Craig one. This is likely an artifact of the small sample size, however both differences are statistically significant.

The reduced-dimension approach identified the top two teams more often (22%) than either win/loss model (18% and 15%, respectively for the Bradley-Terry and Thurstone-Mosteller models) and the inappropriate point-scoring model (7% for the Poisson). However, as one would expect, it performed worse than the true model (79% for the Normal) and the more flexible hybrid formulation (72% for the Annis-Craig).

Finally, the proposed method produced complete rank-orders which were closer to the truth (as measured by Spearman's rank correlation) than either of the misspecified win/loss models, but which were farther removed than both the point scoring methods and the Annis-Craig hybrid. Table 2 summarizes the results of the simulations. Simultaneous Bonferroni 95% confidence intervals for the difference between the reduced-dimension hybrid model and each of the others are given.

Annis: Reduced-Dimension Hybrid Models

95% CI for difference from	Correct #1		Correct Top 2		Rank Correlation	
	Lower	Upper	Lower	Upper	Lower	Upper
Bradley-Terry	0.1190	0.1210	0.0390	0.0410	0.0531	0.0548
Thurstone-Mosteller	0.0691	0.0709	0.0691	0.0709	0.0532	0.0548
Normal	0.0597	0.0603	-0.5703	-0.5697	-0.0133	-0.0128
Poisson	0.2197	0.2203	0.1497	0.1503	-0.0104	-0.0100
Annis-Craig	0.0198	0.0202	-0.5002	-0.4998	-0.0109	-0.0103

Table 2: The reduced-dimension hybrid model outperforms win/loss models when points follow Normal distributions.

3.3 OVERDISPERSED POISSON SIMULATIONS

The third and final simulation generated independent, overdispersed Poisson point totals. The reduced-order hybrid model identified the correct #1 team more often (30%) than either win/loss model (13% each for the Bradley-Terry and Thurstone-Mosteller models) and the misspecified point-scoring model (1% for the Normal). However, the true model (47% for the Poisson) and more flexible hybrid approach (38%) were correct more often.

The top two teams were identified more frequently (16%) by the reduced model than by either win/loss model (7% and 5%, respectively for the Bradley-Terry and Thurstone-Mosteller models) as well as the inappropriate point-scoring model (4% for the Normal). However, as before, the true model (19% for the Poisson) and the Annis-Craig method (29%) were superior.

As was the case with the Normal simulations, the proposed method produced complete rank-orders which were more highly correlated to the truth than either of the misspecified win/loss models, but which were farther removed than either of the point scoring methods or the Annis-Craig hybrid. Table 3 gives simultaneous Bonferroni 95% confidence intervals for the difference between the reduced-dimension hybrid model the others.

3.4 RESULTS

When wins and losses are the only meaningful measure of performance, point-scoring methods essentially chase noise and are therefore undesirable. In this case, the proposed method outperforms both point-scoring models while underperforming the win/loss models and the Annis-Craig hybrid model. It should be emphasized, however, that this simulation is extremely conservative in assessing the reduced model's performance, due to the differences between actual and simulated scoring. This is due to using a mapping function estimated from *actual* scoring data which is at odds with a simulation presuming

95% CI for difference from	Correct #1		Correct Top 2		Rank Correlation	
	Lower	Upper	Lower	Upper	Lower	Upper
Bradley-Terry	0.1690	0.1710	0.0890	0.0910	0.0517	0.0535
Thurstone-Mosteller	0.1691	0.1709	0.1091	0.1109	0.0515	0.0533
Normal	0.2897	0.2903	0.1197	0.1203	-0.0040	-0.0034
Poisson	-0.1703	-0.1697	-0.0303	-0.0297	-0.0119	-0.0114
Annis-Craig	-0.0802	-0.0798	-0.1302	-0.1298	-0.0107	-0.0103

Table 3: The reduced-dimension hybrid model outperforms win/loss models when points follow Overdispersed Poisson distributions.

scores are immaterial. On the other hand, when the true data-generating mechanism is an independent point-scoring process, the reduced-dimension approach outperforms win/loss models and misspecified point-scoring ones. Furthermore, it is comparable to, though slightly worse than, the true model and the Annis-Craig specification.

These results indicate clearly that the proposed method is superior to any of the classical win/loss or point-scoring methods except in the unrealistic situation where the true data-generating mechanism is known in advance. It is robust to model misspecification, and what it sacrifices in performance relative to the Annis-Craig formulation, it gains in computational efficiency and ease of implementation. This point is discussed in detail in the next section.

4 COMPUTATIONAL IMPROVEMENT

The simulations in the previous section show that the reduced-dimensional hybrid model is more robust than either win/loss or point-scoring methods, and therefore desirable in most real situations where the true data-generating mechanism is unknown. Yet it is no surprise that the Annis-Craig model outperforms the simplified hybrid method for two reasons. First, this technique relies on a single run of Annis-Craig model to calibrate the isotonic mapping function, so any lack of fit in the original will be propagated. Second, because it contains half the number of parameters and considers an amalgamation of game data as its response, it is less flexible. Though it lags the complete hybrid approach, this method is substantially easier to implement, runs in a fraction of the time and still offers considerable robustness and flexibility when compared to non-hybrid models.

Table 4 compares average run times (in seconds) and associated standard errors parenthetically for the reduced-dimension hybrid and Annis-Craig models. For each data generating mechanism the proposed method converged dramatically faster and with fewer difficulties than did the original. In all cases,

Annis: Reduced-Dimension Hybrid Models

Convergence Times	Reduced-Dimension	Annis-Craig
Bradley-Terry	5.010 (0.027)	143.382 (1.738)
Normal	4.791 (0.004)	138.792 (2.187)
Overdispersed Poisson	5.139 (0.081)	91.144 (2.071)

Table 4: The reduced-dimension model runs up to 35 times faster than the original.

the difference was at least an order of magnitude. Impressive as they are, these times do not reflect fully the computational improvement of the new technique. In each simulation, the Annis-Craig model was initialized in the neighborhood of the true parameters to help ensure convergence and automate the repeated estimation necessary to conduct the simulations. In practice, when such initial parameter estimates are not known, the complete model may require restarts in different areas of the parameter space before the numerical search settles into a global maximum. (Fortunately, in these circumstances, it is often easy to diagnose whether or not the procedure has converged.)

5 DISCUSSION

A reduced-dimensional variant of the Annis-Craig [1] hybrid ranking algorithm was introduced and examined. Like the original, it is robust to model misspecification and is preferable to pure win/loss or point-scoring methods. The proposed model considers a single quasi-binary outcome, which is a function of both scores, rather than explicitly modeling the binary outcome (win/loss) and both scores hierarchically. Although the additive isotonic mapping function on which it is based requires estimates of the link functions from the Annis-Craig model, it does not require estimates of team strengths, and in practice, this estimation need only be done once, as examination of multiple seasons indicates that the triples $\{E(s_i), E(s_j), \pi_{i,j}\}$ are consistent from year-to-year.

Simulation results suggest that, wherever possible, the Annis-Craig model should be the model of choice for estimating team abilities and constructing rankings. However, in large-data problems – such as ranking all divisions of college football simultaneously – the lower-order hybrid model is a viable alternative. In such a situation, the Annis-Craig model can be fit to a subset of data to obtain estimates of the parameters necessary to fit the isotonic mapping surface. Subsequent estimation based on quasi-binary game credit is achieved in a fraction of the time and will run when the fully specified hybrid approach exhausts the available computational resources.

REFERENCES

- [1] Annis, D. H. and Craig, B. A. (2005) "Hybrid Paired Comparison Analysis, with Applications to the Ranking of College Football Teams" *Journal of Quantitative Analysis in Sports*, 1, Article 3 (<http://www.bepress.com/jqas/vol1/iss1/3>).
- [2] Bacchetti, P. (1989), "Additive Isotonic Models," *Journal of the American Statistical Association*, 84, 289-294.
- [3] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972), "Statistical Inference Under Order Restriction: The Theory and Application of Isotonic Regression," John Wiley (New York).
- [4] Bradley, R. A. and Terry, M. E. (1952), "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39, 324-345.
- [5] Mosteller, F. (1951), "Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations," *Psychometrika*, 16, 3-9.
- [6] Prentice, R. (1976), "A Generalization of the Probit and Logit Methods for Dose Response Curves," *Biometrics*, 32, 761-768.
- [7] Thurstone, L. L. (1927), "A Law of Comparative Judgment," *Psychological Review*, 34, 273-286.