Statistical Practice

Robust Sports Ratings Based on Least Absolute Errors

Gilbert W. BASSETT, Jr.

Rating sports teams based on scores of previous games can be done by estimating the parameters of the standard linear model. Previously proposed estimates have been based on a modified version of least squares in which games with large victory margins were downweighted due to the sensitivity of least squares to outlying observations. This paper considers the least absolute value or L_1 estimate as an alternative method for rating teams. An advantage of L_1 is that it is intrinsically less sensitive to outlying observations, and hence automatically accounts for extreme outcomes. For the ratings version of the linear model it is shown that L_1 is the median of a team's normalized scores and least squares is the average. For teaching purposes the rating problem provides a nonstandard instance of the linear model that students find interesting. Students can tune the estimators to give recent games more weight, and they can compare L_1 to other regression quantiles. Because the estimators can be expressed in terms of the familiar average and median statistics, the ratings context allows insight into how linear model estimators work. Comparison of least squares and L_1 is presented for the 1993 NFL season.

KEY WORDS: Least absolute value; Median; Robust estimation; Ratings.

1. INTRODUCTION

Almost all newspapers publish football and basketball polls, and many, such as *The New York Times* and *USA Today,* publish their own rating systems. In addition to wins and losses, ratings have to account for home-field advantage and the differing quality of opponents. Interest in the ratings is greatest when there is disagreement, especially in cases where there are no playoffs and the national champion is determined by a poll of sportswriters and coaches. For sports like NCAA basketball where there is a championship tournament, the ratings are important for determining who is invited and how teams are seeded. If the game is the thing, the ratings thing would seem a close second.

Rating teams based on past performance has received some attention in the statistics literature. Stefani (1977) described the rating problem, reviewed early sports rating systems, and estimated football ratings using least squares. He reported that in the 1930s there was a widely published P. B. Williamson System that combined calculus and least squares with "hardness of schedule, gameness, and a 'guts' factor." Stefani showed how the rating problem could be posed in terms of the linear regression model, and proposed estimating the ratings by least squares. At about the same time Harville (1977, 1980) constructed ratings for sports teams based on maximum likelihood estimates in which ratings were random variables. Stefani (1980) later showed how the home-field advantage could be incorporated into the ratings model. Following an approach suggested earlier by Leake (1976), the least squares ratings were modified by Stern (1992) to account for the fact that runaway games would unduly affect the least squares estimates. He proposed downweighting large score differences, and produced estimates of the relative strengths of NFL teams. Recently Stern (1995) and Wilson (1995) used least squares to statistically rate college football teams and determine who was number one.

The least squares rating estimates account for home-field advantage and strength-of-schedule. They can be further tuned so that recent performances receive more weight in determining the ratings. This is done by appending a "heteroscedastic" correction to the linear model so that recent games count more heavily in determining the estimates.

In this paper a new method for rating teams is presented. The least absolute error, or L_1 estimator for short, is substituted for least squares as the basic method for rating teams. The motivation for considering the L_1 estimates comes from the fact that they are not distorted by outliers, and hence there is no need for modifications because of runaway scores. Least absolute value estimates are intrinsically less sensitive to extreme values of the dependent variable so that they automatically discount the effect of runaway games. Because of their different sensitivity to the data, the L_1 ratings can be very different from least squares. The potentially large differences will be shown below to correspond to well-known differences that can exist between the average and median statistics.

The connection between L_1 and the median is not surprising. For the simple location problem (where least squares is the sample mean) the L_1 estimate is the sample median. Further, for the p-variable linear model many properties of the L_1 estimate, the $\frac{1}{2}$ regression quantile, mimic median properties; see Koenker and Bassett (1978) and Bassett and Koenker (1978).

Gilbert W. Bassett, Jr. is Professor of Economics and finance at the University of Illinois at Chicago, Chicago, IL 60607-7121 (E-mail: gib@uic.edu).
The author thanks Tom Hettmansperger, Roger Koenker, Jeff Sagarin, Hal Stern, Steve Stigler, and two reviewers for their helpful comments.

For the rating problem the connection between the L_1 estimate and the median, however, is even closer than their common link to the absolute error minimization problem. It is shown that an L_1 rating is actually equal to a median; in particular, it is the median of a team's normalized scores where a normalized score controls for the home-field advantage and opponent strength. This direct link to the median explains why the ratings are insensitive to a few runaway scores.

The connection to the median contrasts with the previously considered least squares ratings. Stefani (1977) had already observed the relation between least squares and normalized scores, but instead of the median, it is the average of scores that controls for home-field advantage and opponent strength. (Normalized scores for the two methods, however, are not identical because each procedure uses its own (different) estimate to account for the home-field advantage and opponent strength.) Properties of the L_1 and least squares ratings therefore follow in a direct and obvious way from their respective characterizations in terms of medians and averages of normalized scores.

The plan for the rest of the paper is as follows. Section 2 presents the rating problem as one of estimating parameters of the usual linear model. A simple example in Section 3 is used to illustrate differences in least squares and L_1 ratings. The example shows how least squares is the average and L_1 is the median of a team's normalized scores; the proof is in the Appendix. Section 4 describes several features of the L_1 ratings including a convenient way to break "ties" when the estimate is not unique. Section 5 compares least squares and L_1 ratings for the 1993 NFL season.

2. THE RATING MODEL

Let teams and games be indexed by $t = 1, \ldots, T$ and $g = 1, \ldots, G$, respectively. Let h_g and a_g denote the home and away teams in game g. Let D_g denote the point difference between the home and away team's score in the gth game, $D_g = S(h_g) - S(a_g)$.

The relative strength of team t is denoted by β_t. The interpretation of the strength coefficients means that $\beta_1 - \beta_2$ measures the difference in the final points scored by teams One and Two if they played at a neutral site. Because *relative* strength depends on rating *differences* (and not the magnitude of the separate β parameters), there is one free parameter for determining rating values. In this paper the Tth team's rating is set to 0 when estimating the rating parameters, and the resulting estimates are then rescaled so that the top-rated team has a value of 100.

The home-field advantage represents the additional points scored by the home team over what it would have scored if the game had been at a neutral site. This home-field parameter is denoted by β_0. (It is assumed that the home-field advantage is identical for all teams; for recent evidence on the home-field advantage see Harville and Smith (1994).)

The difference in the final score can now be expressed by

$$D_g = S(h_g) - S(a_g) = \beta_0 + \beta_{h(g)} - \beta_{a(g)} + e_g \qquad g = 1, \ldots, G. \quad (2.1)$$

This says that the final score is equal to the home-field factor, plus the difference in the ratings of the home and away teams, plus an "error" term. The error can be thought of as the combined effect of all the "breaks" that influence the final score. If team One is at home ($h_g = 1$), team Two is away ($a_g = 2$), team One is rated seven points better than Two ($\beta_1 - \beta_2 = 7$), and $\beta_0 = 3$, then the model predicts that team One will win by ten points. Team Two could win, but it requires that e_g take a large negative value, meaning that the underdog gets the lucky bounces and close calls. (An expanded model for the final scores (rather than the difference in the final scores) that is based on offensive and defensive strength ratings is developed in Bassett (1995).)

To write (2.1) in the form of the standard linear model, let X be a $G \times T$ matrix with elements $\{x_{gt}\}$ where $x_{gt} = 1$ if $t = h(g), x_{gt} = -1$ if $t = a(g)$, and $x_{gt} = 0$ otherwise. Observe that the (parameter) dimension of X is as large as the number of teams, but X is very sparse with mostly zero components. In terms of X Equation (2.1) says

$$D_g = S(h_g) - S(a_g) = \beta_0 + x_{g1}\beta_1 + \cdots + x_{gT}\beta_T + e_g$$

or

$$D_g = S(h_g) - S(a_g) = \beta_0 + \mathbf{x}_g\boldsymbol{\beta} + e_g \qquad g = 1, \ldots, G$$

where $\mathbf{x}_g$ is the vector $(x_{g1}, \ldots, x_{gT})$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)$ is the (column) vector of team ratings.

The least squares ratings are found by minimizing $\sum_{g=1}^{G}(D_g - b_0 - x_g b)^2$. These ratings will be contrasted with the L_1 ratings obtained by minimizing $\sum_{g=1}^{G} |D_g - b_0 - x_g b|$.

3. LEAST SQUARES AS THE AVERAGE AND L_1 AS THE MEDIAN OF NORMALIZED SCORES

Table 1 presents illustrative game results for a league consisting of four teams that have each played five games. It shows, for example, that in the first game the away team (team Two) beat the home team (team One) by a score of 17–14. Notice that games 2 and 5 resulted in very large victory margins of 49 and 35 points for the home team. Standings are shown in Table 2 along with the ratings for each team. The first set of ratings is based on least squares and the second on L_1. For each method there are two ratings

Table 1. Schedule and Outcome of Games

Game	Home	Away	Home score	Away score	Home–away
1	One	Two	14	20	−6
2	Two	Three	38	3	35
3	Four	One	20	13	7
4	Two	Three	17	19	−2
5	One	Three	56	7	49
6	Three	Four	28	20	8
7	Four	One	13	10	3
8	Four	Two	20	10	10
9	One	Three	7	13	−6
10	Two	Four	10	13	−3

Table 2. Standings, Rankings, Ratings

	W	L	OLS rank	OLS rating four = 0	OLS rating max = 100	L_1 rank	L_1 rating four = 0	L_1 rating max = 100
Four	4	1	3	.00	99.66	1	.00	100.00
Three	3	2	4	−9.41	90.25	2	−1.00	99.00
Two	2	3	2	.17	99.83	4	−6.50	93.50
One	1	4	1	.34	100.00	3	−3.50	96.50
Home				6.62			3.50	

reported; the first is standardized so that team four's rating is 0, and the second is an equivalent rescaled version with the top-rated team having a value of 100.

The table shows that least squares and L_1 can produce very different ratings and relative rankings. Team Four is ranked best by L_1, but near the bottom by least squares. Team One is ranked best by least squares, but near the bottom by L_1. There are also differences in the home-field advantage. With least squares the home-field advantage is worth a little more than $6\frac{1}{2}$ points, whereas under L_1 it is only worth $3\frac{1}{2}$ points.

The differences in the ratings reflect the differing sensitivities of least squares and L_1 to the data. Least squares can be strongly influenced by a single observation. In the ratings context a single game with a runaway score affects the ratings of all teams. As described below, a team's rating depends on the relative strength of its opponents. Opponent strength depends in turn on the strength of its opponents, which depends on the strength of opponents' opponents, and so on through connections to all teams. This interdependence combined with the sensitivity of least squares is why only a few lopsided scores greatly influence the ratings estimates.

Dependence on a few observations is the reason for the high ratings for teams One and Two. These teams had one very large victory margin, but otherwise performed poorly; team two only won one other game, and team one lost all its other games. These losing performances do not offset the impact of a single win by a large margin.

This sensitivity of least squares contrasts with L_1, which is not influenced by the games with large margins of victory. The L_1 estimate is not influenced by outlying values of the dependent variable (unless associated with leverage points that, however, are absent from the ratings design matrix). This insensitivity can be understood from the close connection between L_1 and the median that is described below.

It should be emphasized that this illustration does not prove the superiority of one method over another. It is arguable, for example, how much influence a single game should have in determining ratings. What the example shows is that it is possible for the methods to produce very different results.

3.1 Least Squares Ratings and Average Normalized Scores

To understand how the ratings are determined Table 3 presents results for the five games played by team One. The first group of columns shows opponent, final score difference, and the home team. Next is the final score, but adjusted for the home-field advantage. Recall from Table 2 that the home-field advantage is estimated to be worth 6.62 points. Because team One lost its first game by 6 points and the game was at home, it would have lost by $6+6.62 = 12.62$ points if the game had been played at a neutral site. The home-field adjustment is next followed by an accounting for opponent strength. The rating for the opponent, team Two, is .17 (with ratings scaled relative to $\beta_4 = 0$; this means that team two is .17 points better than team Four). Normalizing for both the home-field advantage and opponent strength gives an adjusted score for the first game of $-12.62+.17 = -12.45$; this is an estimate of the difference in the final score if the game had been played at a neutral site against a team with a 0 rating.

The average of all the normalized scores for the games played by team One is seen to be .34. This is identical to team One's overall rating. This is no coincidence: a team's rating is necessarily the average of the normalized scores of all of its games. As demonstrated in the Appendix, the ratings are simultaneously determined so that each team's rating is the average of the normalized scores of its games.

3.2 L_1 Ratings and Median Normalized Scores

To understand how L_1 ratings are determined consider Table 4. As in Table 3 the outcomes of team one's games are listed. The sixth column again is the home-adjusted point difference, but using the L_1 home estimate of 3.50. (The

Table 3. Team One: Analysis of OLS Rating

Game	Opponent	W/L	Point difference	H = +1 A = −1	Home-adjusted difference	Opponent rating	Normalized score
1	Two	L	−6	1	−12.62	.17	−12.45
2	Four	L	−7	−1	−.38	0	−.38
3	Three	W	49	1	42.38	−9.41	32.97
4	Four	L	−3	−1	3.62	0	3.62
5	Three	L	−6	1	−12.62	−9.41	−22.03
				6.62		Ave	.34

Table 4. Team One: Analysis of L_1 Rating

Game	Opponent	W/L	Point difference	H = +1 A = −1	Home-adjusted difference	Opponent rating	Normalized score
1	Two	L	−6	1	−9.50	−6.50	−16.00
2	Four	L	−7	−1	−3.50	.00	−3.50
3	Three	W	49	1	45.50	−1.00	46.50
4	Four	L	−3	−1	.50	.00	.50
5	Three	L	−6	1	−9.50	−1.00	−8.50
				3.50		Median	−3.50

home-field advantage with least squares is the average of the normalized scores of the home teams, whereas for L_1 it is the median.) The home-adjusted score for the first game is therefore $-6.00 - 3.50 = -9.50$. The next column shows opponents' ratings based on the L_1 estimates. The normalized scores are presented in the last column; it shows, for example, that the normalized score for the first game is, $-9.50 - 6.50 = -16.00$.

The median of the normalized scores for all the games played by team One is seen to be −3.50, which is identical to team One's rating. This again is not a coincidence. The L_1 ratings are also determined by normalized scores, but it is the median rather than the average of the normalized scores. This result for L_1 is derived in the Appendix.

Remarks.

1. To compare alternative ratings Stern (1992) considered the number of games in which the ratings correctly predicted the winner; that is, did the favorite (as determined by the ratings) actually win the game? Based on this "within sample" criterion we see that in the above example L_1 outperformed least squares because L_1 predicted seven games correctly (games 3, 4, 5, 6, 7, 8, 10), whereas least squares only had five games correct (games 2, 3, 5, 7, 8).

This summary statistic (number of games correctly predicting the winner) suggests another method for determining ratings, namely, the method defined by the property that it maximize the number of winners. Such a system will not be considered here, except to point out that it is not the same as least squares or L_1. In the above example no ratings can predict more than eight games correctly because games 2 and 4 had the same teams and home field but different winners, and similarly for games 5 and 9. Could any ratings have predicted eight games correctly, and hence on this criterion done better than least squares or L_1? The answer is—yes: set the home advantage to 0 and consider any ratings such that the teams are ranked, Three > Four > Two > One. Such a "maximum number of wins" ranking therefore differs from least squares and L_1. (A modified rating method can be considered by including the "maximum number of wins" as a constraint. For example, define ratings by the property that the sum of absolute errors is minimized, subject to the constraint that the predicted number of winners be as large as possible.)

2. Ratings based on G_w, the number of games up to week w, provide out-of-sample forecasts. These can be used to: (a) assess alternative methods, (b) weight recent performances to achieve better predictions, and (c) compare the ratings against the point spreads. In such a comparison the point spreads are likely to do better because they are based on all current information (such as injured players) that are not reflected in a statistical rating based on only past scores. For comparison of ratings and point spreads see Stern (1992), and for a discussion of point spreads see Bassett (1981).

3. Ratings early in the season require that X_G be of full rank. This rank condition translates to the requirement that there exist a chain of opponents linking all the teams together.

4. The normalized scores can be used for interval and confidence estimates, and to assess model assumptions such as whether variability is the same for all teams. For example, a $Z\%$ empirical rating interval is given by the interval of normalized scores covering $Z\%$ of a team's games.

4. FEATURES OF L_1 RATINGS

The L_1 estimate is computed by solving a linear programming problem; routines for the estimates are available in SAS, S-Plus, STATA, and other statistical packages. The algorithm is a variant of the simplex method modified to take advantage of special features of the L_1 problem; see Koenker and d'Orey (1987, 1993). Many properties of L_1 can be best understood as extended features of the ordinary median and general features of linear programs.

4.1 Multiple Solutions

The L_1 objective function is a polyhedral convex function whose minimum can occur at a vertex; in this case the solution is unique. Alternatively, the minimum can occur on a convex facet of the objective function, in which case there are multiple solutions. In the special case of the median, multiple solutions depend on sample size: an interval of solutions occurs when there is an even number of observations (and the middle order statistics are not equal), and unique solutions occurs when there is an odd number of observations. For the linear model multiple solutions depend on the design, and in general are rare. This is because only a slight perturbation of the design matrix causes the objective function to tilt, thus shifting an otherwise multiple "facet" solution to a unique "vertex" solution.

The design for the ratings model, however, is special and, it turns out, is such that nonunique estimates can occur. One solution is to arbitrarily perturb the design, but this makes the rating estimates depend on an arbitrary tie-breaking rule. A better way to break ties is to weight games so that

Table 5. 1993 NFL Standings and Ratings

	W	L	Points for	Points against	Least squares	L_1
DAL	12	4	376	229	99.99	100
BUF	12	4	329	242	95.15	90
HOU	12	4	368	238	97.53	97.5
NYG	11	5	288	205	94.99	91
KC	11	5	328	291	93.26	93.5
LAA	10	6	306	326	89.69	87.5
SF	10	6	473	295	100	93
DET	10	6	298	292	90.05	90
MIA	9	7	349	351	89.9	84.5
GB	9	7	340	282	93.58	85
MIN	9	7	277	290	90.26	85.5
DEN	9	7	373	284	95.41	94.5
PIT	9	7	308	281	91.85	87
NYJ	8	8	270	247	91.16	85
NO	8	8	317	343	88.99	79.5
SD	8	8	322	290	92.64	96
PHA	8	8	293	315	90.77	85.5
PHX	7	9	326	269	94.29	91.5
CLE	7	9	304	307	89.13	82
CHI	7	9	234	230	90.24	83
SEA	6	10	280	314	89.12	85
ATL	6	10	316	385	85.89	79
LAN	5	11	221	367	81.82	71.5
TB	5	11	237	376	82.81	82
NE	5	11	238	286	86.6	85.5
IND	4	12	189	378	79.08	74
WAS	4	12	230	345	83.82	79
CIN	3	13	187	319	81.91	81.5
Home					2.77	1.50

recent performances receive more influence in determining the estimate.

Consider the modified L_1 problem in which games are weighted by $w_g > 0$,

$$\sum_{g=1}^{G} w_g |D_g - b_0 - x_g b|$$

where $0 \leq w_1 \leq w_2 \leq \cdots \leq w_G$. The unweighted L_1 estimate corresponds to $w_g = 1$, so that early and late season games all receive the same weight. Otherwise, the unequal weights cause more recent performances to receive greater influence in determining the estimate. The weighted estimates can be tuned by adjusting weights and giving greater or lesser influence to early and later games. This weighting makes the estimate depend on recent performance, and moreover it typically results in unique estimates.

To obtain unique ratings with all games nearly equally weighted let weights be $w_g = (1 + \varepsilon g)$, where ε is a very small positive constant and g is the game number. This introduces a very slight bias in favor of recent games that causes the L_1 estimate to be unique.

4.2 The Exact Fit Property

The L_1 estimate corresponds to a hyperplane that fits at least p observations exactly, where p is the number of variables in the design matrix. For the rating design p is equal to the number of teams, minus 1 for the coefficient set to 0, plus 1 for the home advantage. When the solution is unique it will be such that T residuals are zero. If the solution is not unique, then the solution set will be the convex hull of solutions with this exact fit property. In the case of the ordinary median the exact fit property corresponds to the estimate being: 1) equal to one of the observations if it is unique or 2) the convex hull (interval) of observations when the solution is not a single point. This is the "basic solutions" characterization of the solution set to the linear programming problem.

For the ratings case the exact fit property has several implications. The most obvious is that ratings are restricted to a possibly small number of values. Fitted values have the simple form $b_0 + b_{h(g)} - b_{a(g)}$, and the dependent variable (scores) take on integer values. The exact fit property says that the estimated values and integer scores must be equal for at least T observations (games).

As a consequence, L_1 cannot discriminate relative strength well when game scores take on only a few values. Suppose for simplicity that the home factor is 0 and soccer teams are being rated. Let score differences only take integer values between -3 and 3 (negative values corresponding to home team losses), and suppose that there are eight or more teams. Because there are only seven possible score values, there can be only seven allowed rating values so that at least two teams must have identical ratings. This will hold no matter how many games are played. Hence when there are few outcomes, there will be many teams tied with the same rating value. For this reason the ratings are best suited to sports where score differences are more nearly continuous such as football and basketball.

5. RATING NFL TEAMS

Tables 5 and 6 present ratings for NFL teams based on

Table 6. 1993 NFL Team Ranks and Ratings

Team	LS rate	LS rank	L_1 rate	L_1 rank
SF	100	1	93	6
DAL	99.99	2	100	1
HOU	97.53	3	97.5	2
DEN	95.41	4	94.5	4
BUF	95.15	5	90	9.5
NYG	94.99	6	91	8
PHX	94.29	7	91.5	7
GB	93.58	8	85	17
KC	93.26	9	93.5	5
SD	92.64	10	96	3
PIT	91.85	11	87	12
NYJ	91.16	12	85	17
PHA	90.77	13	85.5	14
MIN	90.26	14	85.5	14
CHI	90.24	15	83	20
DET	90.05	16	90	9.5
MIA	89.9	17	84.5	19
LAA	89.69	18	87.5	11
CLE	89.13	19	82	21.5
SEA	89.12	20	85	17
NO	88.99	21	79.5	24
NE	86.6	22	85.5	14
ATL	85.89	23	79	25.5
WAS	83.82	24	79	25.5
TB	82.81	25	82	21.5
CIN	81.91	26	81.5	23
LAN	81.82	27	71.5	28
IND	79.08	28	74	27

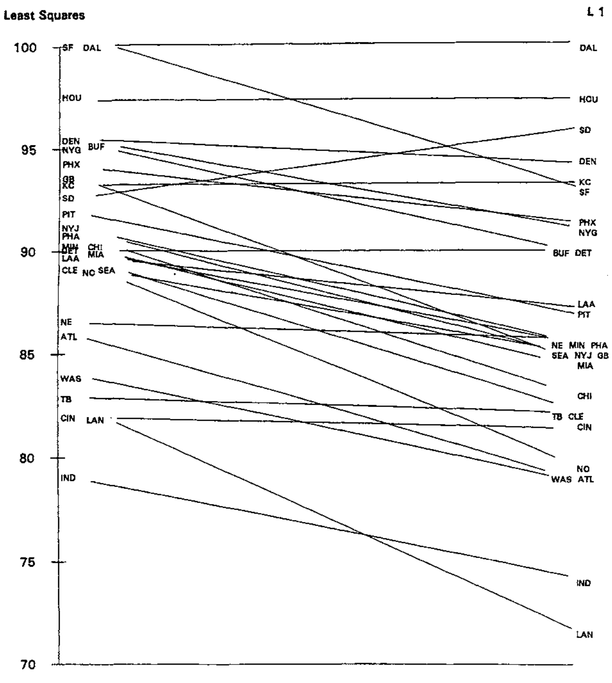

Figure 1. Differences in the 1993 Ratings.

the regular season games of the 1993 season. (The L_1 ratings are the unique estimates for the weighted problem in which games have weight, $1 + .00001^*$week, where week $= 1, \ldots, 18$ is the week of the game.) Table 5 sorts team by their won–lost records, and indicates points scored, points scored by opponents, and the L_1 and least squares ratings. Teams in Table 6 are arranged by least squares ranking, with the L_1 ranking also indicated.

The tables show that the ratings do not always agree with one another; different methods can yield different estimates. For example, San Diego (SD) was ranked third by L_1 and tenth by least squares. On the other hand, least squares had Green Bay (GB), a first-round loser in the playoffs, 8th, whereas L_1 had them only ranked 17th. The table also shows that the methods produce different estimates for the home-field advantage; based on least squares it is worth 2.77 points, but only 1.50 according to L_1. Least squares had Dallas (DAL), the eventual Super Bowl winner, ranked just behind SF, whereas L_1 had Dallas as its the top-rated team, a full 2.5 points better than second-place Houston.

Differences in the ratings are schematically depicted in Figure 1. Teams on the left are displayed according to their least squares ratings and on the right according to L_1 ratings. A line connects the alternative ratings. Flat lines mean similar ratings, and steep lines indicate that the methods produce very different ratings. The steep line connecting San Francisco means that least squares and L_1 had different ratings. (The figure was constructed so that the top-rated teams with both methods are at the same level. Together with the high relative L_1 rating for Dallas, this means that most of the lines to slope down from the least squares side of the figure. An alternative would center the ratings so that they had the same mean or median value (rather than the same maximum value). This would make about half the lines go up and half down, with the Dallas line rising toward the L_1 side of the figure.)

In terms of correct predictions, L_1 outperformed least squares: the higher ranked team won 67% of the games based on least squares, whereas L_1 got 70% of the games correct. For the Super Bowl, least squares made Dallas a five-point favorite, and L_1 had Dallas a ten-point favorite. The Cowboys won 30–13.

6. APPENDIX: PROPOSITIONS

In this Appendix the following propositions are established using the first-order conditions for the estimates.

1. The least squares rating is the *mean* of the least squares normalized scores.
2. The L_1 rating is the *median* of the L_1 normalized scores.

Proof. 1. The first-order condition for the home-field parameter is $\sum_{g=1}^{G}([D_g - x_g\hat{\beta}] - \hat{\beta}_0) = 0$. This means that the home-field advantage is equal to the average of the $D_g - \mathbf{x}_g\hat{\beta}$ values, which is the average victory margin for home teams, after controlling for relative strength. (It is not just the average margin for home teams because strong teams might play a disproportionate number of their games at home or vice versa.)

The first-order condition for the remaining parameters is given by

$$\sum_{g=1}^{G} (D_g - \hat{\beta}_0 - x_g\hat{\beta})x_g = 0.$$

The individual components in this vector equality are

$$\sum_{g=1}^{G} (D_g - \hat{\beta}_0 - x_g\hat{\beta})x_{gt} = 0 \qquad t = 1, \ldots, T-1$$

(there are $T - 1$ variables because β_T is set to 0) or on restricting to the summands with $x_{gt} \neq 0$,

$$\sum_{g:x_{gt}=\pm 1} (D_g - \hat{\beta}_0 - x_g\hat{\beta})x_{gt}$$

$$= \sum_{g:x_{gt}=\pm 1} (x_{gt}D_g - x_{gt}\hat{\beta}_0 - x_{gt}x_g\hat{\beta}) = 0$$

$$t = 1, \ldots, T-1.$$

Because each term in the sum only depends on games played by team t, the equation can be most easily represented in terms of the following "team" notation.

Let N_t denote the number of games played by team t. Let $\text{Opp}(t_i)$ denote the opponent of team t in its ith game, and let $h(t_i)$ be a home indicator: $+1$ if team t is at home in its ith game and -1 if it is away. Let $S(t_i)$ be the points scored by team t in its ith game, $S(\text{Opp}(t_i))$ the points scored by the opponent, and $\delta t_i = S(t_i) - S(\text{Opp}(t_i))$ the difference in the final score.

Now observe that when $x_{gt} = 0$, the terms in the summand become

1. $x_{gt}D_g = \delta t_i = S(t_i) - S(\text{Opp}(t_i))$

2. $x_{gt}\hat{\beta}_0 = h(t_i)\hat{\beta}_0$

3. $x_{gt}\mathbf{x}_g\hat{\boldsymbol{\beta}} = \hat{\beta}_t - \hat{\beta}_{\text{Opp}(ti)}$. The first-order conditions therefore translate into

$$\sum_{i=1}^{N_t} (\delta t_i - h(t_i)\hat{\beta}_0 + \hat{\beta}_{\text{Opp}(t_i)} - \hat{\beta}_t) = 0 \qquad t = 1, \ldots, T-1$$

or on setting $\text{Norm}(t_i) = \delta t_i - h(t_i)\hat{\beta}_0 + \hat{\beta}_{\text{Opp}(ti)}$, which is the normalized score in the ith game played by team t,

$$\sum_{i=1}^{N_t} (\text{Norm}(t_i) - \hat{\beta}_t) = 0 \qquad t = 1, \ldots, T-1.$$

This means that the estimate is the average of team t's normalized scores.

2. The proof for L_1 is similar. The first-order condition for the home parameter says that the estimates β_0^* and $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_{T-1}^*)$ satisfy

$$\sum_{g=1}^{G} \text{sgn}^*(D_g - x_g\beta^* - \beta_0^*) = 0.$$

(sgn^* is a shorthand for the subdifferential of the absolute value function; $\text{sgn}^*(v) = 1$ or -1 as $v > 0$ or $v < 0$ and $\text{sgn}^*(0)$ is the interval $[-1, +1]$. The summation means that the (possibly interval-valued) summands are added (setwise), and the equality condition means that zero is an element of the possibly interval-valued term on the left-hand side.) The important thing to note in this is that the median of the values $(Z_1, \ldots, Z_n)$ is any z' such that $\sum \text{sgn}^*(Z_i - z') = 0$; see Rockafellar (1970) for an analysis of convex functions using subdifferentials. The condition therefore says that the home-field advantage is the median of $D_g - \mathbf{x}_g\boldsymbol{\beta}^*$.

The first-order condition for the remaining parameters is given by

$$\sum_{g=1}^{G} \text{sgn}^*(D_g - \beta_0^* - x_g\beta^*)x_g = 0.$$

The individual components of the vector equality are

$$\sum_{g:x_{gt}=\pm 1} \text{sgn}^*(D_g - \beta_0^* - x_g\beta^*)x_{gt}$$

$$= \sum_{g:x_{gt}=\pm 1} \text{sgn}^*(x_{gt}D_g - x_{gt}\beta_0^* - x_{gt}x_g\beta^*) = 0$$

$$t = 1, \ldots, T-1.$$

Translating again into team notation gives

$$\sum_{i=1}^{N_t} \text{sgn}^*(\delta t_i - h(t_i)\beta_0^* + \beta_{\text{Opp}(t_i)}^* - \beta_t^*) = 0 \qquad t = 1, \ldots, T-1$$

or, with normalized scores being defined as above except that L_1 replaces least squares,

$$\sum_{i=1}^{N_t} \text{sgn}^*(\text{Norm}(t_i) - \beta_t^*) = 0 \qquad t = 1, \ldots, T-1$$

which says that the L_1 estimate is the median of team t's normalized scores.

[Received June 1994. Revised June 1996.]

REFERENCES

Bassett, G. W. (1981), "Point Spreads vs. Odds," *Journal of Political Economy*, 80(4), 752–768.

——— (1995), "Predicting the Final Score," manuscript.

Bassett, G. W., and Koenker, R. (1978), "The Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association*, 73(363), 618–622.

Harville, D. (1977), "The Use of Linear-Model Methodology to Rate High School or College Football Teams," *Journal of the American Statistical Association*, 72, 278–289.

——— (1980), "Predictions for National Football League Games with Linear-Model Methodology," *Journal of the American Statistical Association*, 75, 516–524.

Harville, D. A., and Smith, M. H. (1994), "The Home-Court Advantage: How Large and Does it Vary from Team to Team," *The American Statistician*, 48(1), 22–28.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

Koenker, R., and Bassett, G. W., Jr. (1978), "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and d'Orey, V. (1987, 1993), "Computing Regression Quantiles," *Applied Statistics*, 36, 383–393; 43, 410–414.

Leake, R. J. (1976), "A Method for Ranking Teams with an Application to 1974 College Football," *Management Science in Sports*, North Holland.

Rockafellar, R. T. (1970), *Convex Analysis*, Princeton, NJ: Princeton University Press.

Stefani, R. T. (1977), "Football and Basketball Predictions Using Least Squares," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7, 117–121.

——— (1980), "Improved Least Squares Football, Basketball, and Soccer Predictions," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-10(2), 116–123.

Stern, H. (1992), "Who's Number One?—Rating Football Teams," in *Proceedings of the Section on Statistics in Sports* 1992, pp. 1–6.

——— (1995), "Who's Number 1 in College Football? ... and How Might We Decide?," *Chance*, 8(3), 7–14.

Wilson, R. L. (1995), "The 'Real' Mythical College Football Champion," *OR/MS Today*, 24–29.

LINKED CITATIONS

Robust Sports Ratings Based on Least Absolute Errors
Gilbert W. Bassett, Jr.
The American Statistician, Vol. 51, No. 2. (May, 1997), pp. 99-105.
Stable URL:
http://links.jstor.org/sici?sici=0003-1305%28199705%2951%3A2%3C99%3ARSRBOL%3E2.0.CO%3B2-A

References

Point Spreads versus Odds
Gilbert W. Bassett, Jr.
The Journal of Political Economy, Vol. 89, No. 4. (Aug., 1981), pp. 752-768.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198108%2989%3A4%3C752%3APSVO%3E2.0.CO%3B2-U

Asymptotic Theory of Least Absolute Error Regression
Gilbert Bassett, Jr.; Roger Koenker
Journal of the American Statistical Association, Vol. 73, No. 363. (Sep., 1978), pp. 618-622.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197809%2973%3A363%3C618%3AATOLAE%3E2.0.CO%3B2-3

The Use of Linear-Model Methodology to Rate High School or College Football Teams
David Harville
Journal of the American Statistical Association, Vol. 72, No. 358. (Jun., 1977), pp. 278-289.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197706%2972%3A358%3C278%3ATUOLMT%3E2.0.CO%3B2-W

Predictions for National Football League Games Via Linear-Model Methodology
David Harville
Journal of the American Statistical Association, Vol. 75, No. 371. (Sep., 1980), pp. 516-524.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198009%2975%3A371%3C516%3APFNFLG%3E2.0.CO%3B2-I

LINKED CITATIONS

The Home-Court Advantage: How Large Is It, and Does It Vary from Team to Team?
David A. Harville; Michael H. Smith
The American Statistician, Vol. 48, No. 1. (Feb., 1994), pp. 22-28.
Stable URL:
http://links.jstor.org/sici?sici=0003-1305%28199402%2948%3A1%3C22%3ATHAHLI%3E2.0.CO%3B2-W

Regression Quantiles
Roger Koenker; Gilbert Bassett, Jr.
Econometrica, Vol. 46, No. 1. (Jan., 1978), pp. 33-50.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197801%2946%3A1%3C33%3ARQ%3E2.0.CO%3B2-J

Algorithm AS 229: Computing Regression Quantiles
Roger W. Koenker; Vasco D'Orey
Applied Statistics, Vol. 36, No. 3. (1987), pp. 383-393.
Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281987%2936%3A3%3C383%3AAA2CRQ%3E2.0.CO%3B2-8