Statistical Modelling 2007; **7**(4): 301–313

Quantile regression for rating teams

Gilbert W Bassett Jr
Department of Finance, University of Illinois at Chicago, US

Abstract: Quantile regression is proposed for modeling game outcomes and as the basis for rating teams. The model includes the standard location model for team strength as a special case, while allowing for a richer specification in which teams differ according to the quantiles of the outcome distribution. Team ratings are defined as the handicap needed to equalize the outcome of a contest. With teams differing by quantiles, this leads to a class of ratings that depend on where in the outcome distribution the outcome is equalized. Relationships with betting games are discussed. The approach is illustrated by rating National Football League (NFL) teams based on game results for the 2005 season.

Key words: handicaps; odds; point spreads; quantile regression; sports ratings

Received August 2006; **revised** March 2007; **accepted** May 2007

1 Introduction

What does it mean when Team A is ranked number one and Team B is ranked number two? One interpretation is that if A and B were to play on a neutral site (no home-field advantage) then A would on *average* win by $s > 0$ points. To make the contest evenly matched requires giving s points to the lower ranked Team B. If D_{AB} denotes A's score minus B's score, then 'A better than B' if $E(D_{AB^{-s}}) = 0$ for some $s > 0$.

Alternatively, 'A better than B' could mean that A has a better than 50–50 chance of beating B. Instead of being about the average victory margin, this interpretation is about who is more likely to win the game. Expressed in handicap terms it means giving points to B in order to make the handicap-adjusted outcome 50–50; 'A better than B' if there is a $z > 0$ such that $Pr(D_{AB^{-z}} > 0) = Pr(D_{AB^{-z}} < 0) = 0.5$.[1]

Both these interpretations (and others) are reasonable. But they are different—expected and median values of a random variable need not agree—and can lead to different ratings and rankings. This means that the ranking depends on how 'A better

Address for correspondence: Department of Finance, University of Illinois at Chicago, 601 South Morgan (MC 168), Chicago, Illinois 60607–7121. E-mail: gib@uic.edu

[1]For ease of exposition here and in the following, we ignore discrete, integer, scoring, and suppose D to have a continuous range of outcomes.

 10.1177/1471082X0700700402

than B' is interpreted. We can have 'A better than B' based on the expected outcome, while at the same time B is more likely to win the game.

For many of the standard models the alternative ranking interpretations are restricted to be equivalent. For example, consider the linear model, $Pr(D_{AB} < z) = F_{\text{o}}(z - (\beta_{\text{A}} - \beta_{\text{B}}))$ where F_{o} is a symmetric-about-zero probability distribution and the βs are relative strength parameters for each team. For the paired comparison case, observations are indicators, $I[D_{AB} > 0]$, (Team A or B wins the contest) and the probability that A beats B is $Pr(D_{AB} > z) = F_{\text{o}}(\beta_{\text{A}} - \beta_{\text{B}})$; see David (1988: 7). The Thurstone-Mosteller (normal) and Bradley-Terry (Zermelo) models correspond to alternative specifications of F_o. In this case the different rankings are all identical and correspond to the ordering of the βs—because the model only allows teams to differ by a location parameter. It follows that 'A better than B' at the expected value implies 'A better than B' at the median as well as at every other part of the distribution.

This contrasts with the quantile model considered here. The model allows for a richer description of the outcomes when A plays B so that alternative ranking criteria can lead to different rankings. Compared to the standard approach where teams only differ by location, the quantile model expresses relationships that can vary depending on the quantiles of the distribution. It means that the ratings are allowed to depend on how, 'A better than B' is interpreted.

In this paper, ratings are defined as the handicap needed to make handicap-adjusted outcomes evenly matched. However, as suggested earlier 'equally matched' depends on which property of the outcomes is equalized. The handicap needed to equalize the expected scores of the two teams can differ from what is needed to equalize the chances of winning. The usefulness of the quantile model is that it can express these different versions of 'equally matched'. Hence, ratings and rankings ultimately depend on how 'A better than B' is interpreted.

The ratings are illustrated for the 32 NFL teams based on outcomes of all 2005 regular season games. The ratings are compared at the mean as well as at other parts of the outcome distribution. Interest is in assessing differences in the ratings and rankings according to the different criteria.

Ratings for the paired comparison model are based on win/loss outcomes. For the history of paired comparison models and approaches see David (1988:11–14). Harville (1977) was the first to propose the linear model for rating sports teams based on point differentials. For more recent discussions of sports rating issues see Stern (1995, 2004), and Frey (2005). The least absolute value estimator or L1 estimator, which is the 0.5 regression quantile, was proposed for rating teams in Bassett (1997), but the motivation was robustness, rather than ratings at alternative parts of a distribution. For comprehensive review of general quantile modeling and estimation see Koenker (2005), Koenker and Bassett (1978), and Koenker and Hallock (2001).

The next section introduces the quantile model and defines the ratings in terms of point-equalizing handicaps. Section 3 discusses relationships between the ratings and different wagering games. Section 4 discusses estimation issues including the home field advantage and incorporating constraints due to the nature of the paired comparison data. Section 5 presents ratings for the NFL teams based on the alternative rating criteria. While the alternative rating criteria sometimes lead to similar rankings, there are also cases where there are large differences that depend on where comparisons are made in the outcome distribution.

2 Teams, leagues, handicaps

2.1 The quantile model

A league consists of T teams, 1, ..., T. A game is a contest between a pair of teams. To present the ranking and rating ideas, we initially suppose that there is no home-field advantage, as if all the contests are at a neutral site.

The outcome when i plays j is a random variable denoted by $D_{ij} = S_i - S_j$ team i's score minus j's score (so $D_{ji} = -D_{ij}$).[2] The distribution and quantile functions of the D_{ij} are denoted by F_{ij} and Q_{ij}, respectively.

As noted earlier, the outcome distributions with the standard linear model are identical except for location. That is, $D_{ij} = \beta_i - \beta_j + \varepsilon$, and the quantile functions are $Q_{ij}(\theta) = Q_o(\theta) + \beta_i - \beta_j$, where $Q_o(\theta)$ is the symmetric[3] quantile function of ε. In this case the β_i are relative strength parameters that location-shift the outcome distribution ('relative' because the rating values are determined up to an additive constant).

The quantile model generalizes the standard approach by allowing team effects to vary depending on the quantile of the distribution; the quantile function of D_{ij}, $i < j$, is

$$Q_{ij}(\theta) = Q_o(\theta) + \beta_i(\theta) - \beta_j(\theta)$$

where Q_0 is a symmetric (around zero) quantile function.

To understand what the model says about outcomes, first notice that $Q_o(\theta)$ is the quantile function for two evenly-matched teams, ($\beta_i(\theta) - \beta_j(\theta) = 0$, *all* θ). If the contest is instead between team i and team j where $\beta_j(\theta) = 0$, all θ then $\beta_i(\theta)$ represents how team i alters outcomes relative to what would have happened in an

[2]Let S_{ij} be the total points scored when i plays j. In this paper we restrict attention to ratings and rankings based on the differential, $D_{ij} = S_{ij} - S_{ji}$, and leave to future research the modeling and estimation of S_{ij}. The model for S_{ij} leads to separate offensive and defensive ratings (points scored by i's offense against j's defense) as well as an estimate of the total points scored by both teams, the so-called over/under.

[3]Since $D_{ji} = -D_{ij}$ the quantile functions for $j > i$ satisfy, $Q_{ji}(\theta) = -Q_{ij}(1 - \theta)$, which means $Q_o(\theta) = -Q_o(1 - \theta)$.

evenly-matched contest. If $\beta_i(\theta) > 0$ the outcomes are shifted in favor of team i. In the special case where $\beta_i(\theta) = \beta_i$, a constant for all θ—the standard model—the distribution is distorted by the same β_i at all parts of the distribution. In the general case, however, the distortion is θ-dependent, larger or smaller depending on the magnitude of $\beta_i(\theta) > 0$.

To express the outcome D_{ij} so that it has quantile function Q_{ij}, recall that the random variable $D = Q(U), U$ uniform, has quantile function, Q. Hence, $D_{ij} = Q_0(U) + \beta_i(U) - \beta_j(U)$, is the random variable with the quantile function Q_{ij}. (Note that the same U determines $Q_0(U), \beta_i(U)$ and $\beta_i(U)$—they are comonotonic; see Koenker (2005) for connections between comonotonicity and regression quantile.)

Remarks: (1) The $\beta_i(\theta), i = 1, \ldots, T$ themselves are not quantile functions. For one thing they are not restricted to be monotone in θ. They can be viewed as the amount needed to turn an evenly-matched contest into one between unevenly-matched teams. (2) The $\beta_i(\theta)$ are not arbitrary. Since $Q_{ij}(\theta)$ is a quantile function, it is increasing, and hence the $\beta_i(\theta)$ satisfy, $q_{ij}(\theta) = q_o(\theta) + \frac{d\beta_i(\theta)}{d\theta} - \frac{d\beta_j(\theta)}{d\theta} > 0$, where $q_{ij}(\theta) = f(Q(\theta))^{-1}$ denotes the derivative of Q_{ij}.

2.2 Ratings

Team *ratings* are numerical values denoted by $\delta_1, \ldots, \delta_T$. The *rankings* are the ordinals obtained from the rating vector with the top rated team ranked '1', to the lowest rated team (assuming no ties) with rank 'T'.

The δ_i will be defined so that handicap adjusted outcomes result is an evenly matched contest. Let HD_{ij} denote the handicap adjusted outcome when i plays j, $HD_{ij} = D_{ij} - (\delta_i - \delta_j)$. The $\delta_i - \delta_j$ indicate how the score must be adjusted to equalize the outcome. Except in special cases (which correspond to the standard model) the handicap varies according to the outcome-property to be equalized.

The ratings that equalize the probability of winning are denoted by $\delta_i(0)$. They are defined so that the handicap-adjusted probability that i beats j is the same as the probability that j beats i, $Pr[HD_{ij} > 0] = Pr[HD_{ji} > 0]$, or $1 - F_{ij}(\delta_i(0) - \delta_j(0)) = F_{ij}(\delta_i(0) - \delta_j(0))$. This says that $\delta_i(0) - \delta_j(0)$ is the median of D_{ij}. For the quantile model where the median of D_{ij} is $\beta_i(0.5) - \beta_j(0.5)$ the team ratings therefore are, $\delta_i(0) = \beta_i(0.5), i = 1, \ldots, T$.

The ratings, $\delta_i(0)$, will correspond to the $s = 0$ special case of ratings denoted by $\delta_i(s)$. Select a value for $s, s > 0$ and consider the events: (a) i beats j by more than s points; (b) j beats i by more than s points. We want to select handicaps so that these events have the same probability. That is, $Pr[HD_{ij} > s] = Pr[HD_{ji} > s]$, or, $1 - F_{ij}(\delta_i(s) - \delta_j(s) + s) = F_{ij}(\delta_i(s) - \delta_j(s) - s)$. Expressed in terms of the quantile functions this says, $\delta_i(s) - \delta_j(s) = 1/2[Q_{ij}(\theta) + Q_{ij}(1-\theta)]$, where, $s = 0.5[Q_{ij}(1-\theta) - Q_{ij}(\theta)]$.

For the standard model in which the $\beta(\theta)$ do not depend on θ this means,

$$\delta_i(s) - \delta_j(s) = 0.5\left[Q_{ij}(\theta) + Q_{ij}(1-\theta)\right] = 0.5\left[Q_o(\theta) + \beta_i - \beta_j + Q_o(1-\theta) + \beta_i - \beta_j\right] = \beta_i - \beta_j$$

so that $\delta_i(s) = \beta_i$ for all s, that is, the same rating holds for any s.

For the quantile model, in contrast, the ratings depend on where in the distribution outcomes are equalized. That is,

$$\begin{aligned}\delta_i(s) - \delta_j(s) &= 0.5[Q_{ij}(\theta) + Q_{ij}(1-\theta)] = 0.5[Q_o(\theta) + \beta_i(\theta) - \beta_j(\theta) \\ &\quad + Q_o(1-\theta) + \beta_i(1-\theta) - \beta_j(1-\theta)] \\ &= 0.5[\{\beta_i(\theta) + \beta_i(1-\theta)\} - \{\beta_j(\theta) + \beta_j(1-\theta)\}]\end{aligned}$$

So that the rating for team i is $\delta_i(s) = 1/2[\beta_i(\theta) + \beta_i(1-\theta)]$.

Notice that the score-parameterized ratings can be alternatively represented by the probability values they equate. Instead of $\delta_i(s)$ we could write $\delta_i(\theta)$. Depending on the context we will variously refer to the ratings by $\delta_i(s)$ or $\delta_i(\theta)$.

Finally, the ratings that equalize *expected* outcomes are denoted by δ_i^E. These are defined so that, $E(HD_{ij}) = 0$, all i, j, or, $E(D_{ij}) - (\delta_i^E - \delta_j^E) = 0$. For the general quantile model this means δ_i^E is the integral over all the quantile ratings, θ in (0, 1) (that is, $\int_0^1 Q(\theta)d\theta = \int x dF(x)$). This is trivially also true in the special case of the standard model where all the quantile effects are identical.

3 Point spreads, odds, ratings

This section examines the relationship between ratings and various wagering games such as point spreads and odds. In an earlier comparison, Bassett (1981), one of the questions was how point spread betting (a wager at even odds on the point spread adjusted winner) would map to odds betting (a wager on the outright winner but at variable odds). We first consider a general wagering game whose equilibrium values trace out the consensus beliefs/probability distribution about the contest. When this distribution satisfies the location-shift model, there is a one-to-one correspondence between point spreads and odds. The correspondence, however, fails if beliefs follow the quantile model. It means that for teams i and j, a point spread of 7 points can correspond to an odds of 3:1, whereas the same spread of 7 points for teams i and k can translate to an odds of 4:1. The section concludes with a description of the betting game that gives rise to the ratings as handicaps.

Consider wagers, depending on s and θ, presented in Table 1.

There are two possible wagers: an 'i' bet on the event $D_{ij} - s > 0$ ('i' beats the spread) or a 'j' bet $D_{ij} - s < 0$ ('j' beats the spread). For each outcome the net

Table 1 Point spread and odds wagers

	Event: team wins	Handicap adjusted event	Wager $(s:\theta)$ i bet: $(D_{ij} - s > 0)$	 j bet: $(D_{ij} - s < 0)$
'i' bet	$S_i - S_j = D_{ij} > 0$	$D_{ij} - s > 0$	$\theta/(1+\theta)$	$-\theta/(1+\theta)$
'j' bet	$S_i - S_j = D_{ij} < 0$	$D_{ij} - s < 0$	-1	1
		$D_{ij-s} = 0$	0	0

payoff is presented in the table. The wager on 'i' pays \$ $\theta/(1+\theta)$ in case $D_{ij} - s > 0$, but loses \$ 1 if $D_{ij} - s < 0$ (there is no change in wealth when $D_{ij} - s = 0$. In a similar fashion the table shows the payoffs for a wager on the other team, 'j'. The standard point spread betting game corresponds to $(s, 0.5)$, while the odds wagering game is $(0, \theta)$.

An equilibrium (s, θ) is defined so that a bookie's profits are perfectly hedged and do not depend on the outcome of D_{ij}. Solving for the equilibrium s, θ gives: $\{(s,\theta)| Pr(j : s,\theta) = \theta\}$ where $Pr(j : s, \theta)$ denotes the proportion of 'j' bets in the market.[4] If, in addition, it is assumed that the proportions adjust so as to eliminate wagers with positive expected value then the equilibrium set will be the graph of the distribution or quantile function, $\{s,\theta|\theta = F_{ij}(s)\} = \{s,\theta|s = Q_{ij}(\theta)\}$.

For the standard model, where the distribution of D_{ij} is $F_0(\beta_i - \beta_j)$ the equilibrium values are given by $\{s,\theta|\theta = F_0(\beta_i - \beta_j)\} = \{s,\theta|s = Q_0(\beta_i - \beta_j)\}$. The point spread game (at even odds) has point spread, $\beta_i - \beta_j$. The odds game on the winner of the contest ($s = 0$) has odds of $\theta = F_0(\beta_i - \beta_j)$. We see that there is a one-to-one relationship between the odds and the point spread with the odds $\theta = F_0(\beta_i - \beta_j)$ and the point spread $s = Q_0(\theta)$.

For the quantile model, however, there need not be a mapping from *the* point spread to *the* odds that is independent of the teams. The point spread when i plays j is given by, $s_{ij} = \beta_i(0.5) - \beta_j(0.5)$, whereas the odds are determined so that, $F_{ij}(0) = \theta_{ij}$, where $Q_{ij}(\theta) = Q_0(\theta) + \beta_i(\theta_0) - \beta_j(\theta_0)$. A point spread of 7 points for i and j can correspond to odds of 3:1, whereas the same spread of 7 for i and k means odds of 4:1.

Finally, we identify the wagering game whose equilibrium corresponds to the handicaps that generate the rating values defined in the previous section. In this game each wager is a point spread bet (at 1:1 odds), but instead of being on the (point spread adjusted) event that a team wins the game, it is on the (point spread adjusted) event in which a team wins by at least z points.

[4] That is, let N denote the total number of wagers, $Pr(i{:}s, \theta)$ the proportion of bets on 'i', and $1 - Pr(i{:}s, \theta)$ the proportion of bets on 'j'. If the outcome is $D_{ij} - s > 0$ the bookie's profits are, $N(-[\theta/(1+\theta)](1 - Pr(i{:}s, \theta) + Pr(i{:}s, \theta)))$. If, instead, the outcome is $D_{ij} - s < 0$, profits are $N([0/(1+0](1 - Pr(i{:}s, \theta) - Pr(i{:}s, \theta)))$. Solving gives the equilibrium.

Table 2 Point spread wagers on winning by more than z points

	Event: team wins by $z \geq 0$ points	Handicap adjusted	Wager $s(z)$	
			i bet: $(D_{ij} - s > z)$	j bet: $(D_{ij} - s < -z)$
'i' bet	$S_i - S_j = D_{ij} > z$	$D - S > z$	1	−1
'j' bet	$S_i - S_j = D_{ij} < -z$	$D - S < -z$	−1	1
	$-z < D_{ij} < z$	$z < D - S > z$	0	0

For a given $z > 0$, the game is defined in Table 2. Wagering is on the team that wins by at least z points where z is pre-specified. The point spread s is now set to equalize the proportion of bets on a team winning by z or more points.

Let $Pr(\text{'}i\text{'}{:}z, s)$ denote the proportion of wagers on team i given z and s, so that $1 - Pr(\text{'}j\text{'}{:}z, s) = 1 - Pr(\text{'}i\text{'}{:}z, s)$ is the proportion of bets on team j. Solving as earlier, we find that the bookie is hedged by setting s so that the proportion of bets on i and j is equal. Further, in order that neither wager has positive expected value gives $Pr\{\text{'}i\text{'}{:}z, s\} = Pr\{D_{ij} - s > z\} = 1 - F_{ij}(z + s)$. Finally, if the D_{ij} follow the quantile model then the equilibrium condition, $1 - F_{ij}(z + s) = F_{ij}(s - z)$ is satisfied by $s = \delta_i(z) - \delta_j(z)$ where $\delta_i(z) = 0.5[\beta_i(\theta_0) + \beta_i(1 - \theta_0)]$. This identifies the wagering game that generates the ratings.

4 Estimation: home-field, symmetry, weighting

Ratings are estimated based on outcomes of the 256 regular season games in the 2005 NFL regular season.[5] The teams in contest $g = 1, \ldots, 256$, are $i(g)$, $j(g)$, $i < j$. Each contest has a home team, $h(g)$, and an away team $a(g)$. We assume there is a home-field advantage denoted by h_0, which does not depend on the teams and is the same at each home site. The home field shifts expected outcomes in favor of the home team by h_0 over what would have happened at a neutral site. The $D_g = D_{i(g)j(g)}$, $i < j$, are independent with quantile function,

$$Q_g(\theta) = Q_{i(g)j(g)}(\theta) = h_0 I[i(g) = h(g)] - h_0 I[i(g) \\ = a(g)] + Q_o(\theta) + \beta_{h(g)}(\theta) - \beta_{a(g)}(\theta)$$

where $Q_0(\theta)$ is symmetric about zero.

4.1 Home-field

The home-field parameter is estimated from the conditional expectation model using ordinary least squares. The conditional expectation model is:

[5] Data was obtained from Kenneth Massey (2006) website.

$$E\left(D_{h(g),a(g)}\right) = h_0 + \beta^E_{h(g)} - \beta^E_{a(g)}.$$

The estimate for the home-field advantage in 2006 is 3.64. The estimates for β_i^E are presented below after discussing the quantile estimates.

The quantile coefficients are estimated after first adjusting the outcomes by the home-field advantage, $D_g = D_{i(g)j(g)} - 3.64$, which has quantile function,

$$Q_{ij}(\theta) = Q_o(\theta) + \beta_i(\theta) - \beta_j(\theta).$$

4.2 Symmetry

The symmetry constraint, $Q_0(\theta) = -Q_0(1-\theta)$ is imposed on the estimates by jointly estimating the θ and $1-\theta$ regression quantiles. This is done by solving the 'stacked' problem,

$$[\hat{Q}_o(\theta), \hat{\beta}(\theta) : -\hat{Q}_o(\theta), \hat{\beta}(1-\theta)] = \arg\min_{(a,b_1),(a,b_2)}\{S_\theta(a, b_1 : D, X) + S_{1-\theta}(-a, b_2 : D, X)\}$$

where $S_\theta(a, b : D, X)$ is the criterion function for the θth regression quantile, D is the vector of team scores and X is the associated design.[6] This imposes symmetry on the intercept while allowing the quantile team coefficients to be separately determined at θ and $1-\theta$. (The minimization problem (in θ and $1-\theta$) can be reduced to one in only θ by noting that for the special team design, $S_{1-\theta}(-a, b_2 : D, X) = S_\theta(a, b_2 : -D, -X)$).

4.3 Weighting

The quantile regression estimates are not guaranteed to be unique because they solve a *polyhedral* convex minimization problem. This non-uniqueness is the linear model version of 'jumps' in the ordinary sample quantile function, or the interval of medians that occurs with an even number of observations. For general designs, non-uniqueness is not important because it is rare and, if it does occur, unique estimates are achieved with an epsilon perturbation of the design.

For the paired comparison design, however, non-unique estimates tend to be the rule. To make the estimates unique with an arbitrary, though slight, change in the design is not reasonable because it introduces arbitrariness into the ratings and rankings. To resolve non-uniqueness we propose instead (similar to what was done with the L-one estimate [see Bassett 1997:102]) to use a weighted estimate that puts slightly more weight on more recent observations. The resulting weighted

[6] See Koenker (1984) who uses stacking for testing constrained and unconstrained models.

estimates are unique and plausibly attribute more information about current team strength to more recent performances.[7] For the estimates reported below outcomes are weighted by $(1 + 0.0001 * \text{week})$ where week refers to the week of the game. The first week of the season is week '1' while the final week with the most recent outcome is week '17'.

Remark: We know that $D_{ij} = -D_{ji}$ so that $E(D_{ij}) = -E(D_{ji}) = -E(-D_{ij})$. Similarly, $\text{Median}(D_{ij}) = -\text{Median}(D_{ji}) = -\text{Median}(-D_{ij})$. But the quantiles do not satisfy these equivariance properties. The quantiles for 'ij' and 'ji' are related by $Q_{ij}(\theta) = -Q_{ji}(1-\theta)$ so that θth quantile, $(\theta \neq 0.5)$ of D_{ij} will generally not be the same as minus the θ quantile of D_{ji}. To say the same thing, D_{ij} and $-D_{ij}$ do not have the same quantile functions. An implication is that for quantile estimation (with $\theta \neq 0.5$) one has to be careful about data entry when the same two teams play multiple contests against each other (as occurs with the NFL schedule). To fix ideas, let the contestants for games 1 and 2 be the same two teams, 5 and 6: $i(1), j(1) = (5, 6)$ and $i(2), j(2) = (5, 6)$. Suppose the outcomes of the games are $D_1 = 7$ and $D_2 = -10$, which means team '5' won the first game by 7 points while team '6' won the second game by 10 points. We could write the first two outcomes as $(7, -10)$ with a design, $\beta_5 - \beta_6$ for the first outcome and $\beta_5 - \beta_6$ for the second outcome. Or, we could write the outcome as (7, 10) with $\beta_5 - \beta_6$ for the first outcome and $\beta_6 - \beta_5$ the second outcome. Either of these works when estimating the expected value or median (and least squares and the 0.5 regression quantile, the least absolute error estimator, are invariant to either specification). To estimate the θth quantile however we cannot have $S_5 - S_6 = 7$ for game one, and $S_6 - S_5 = 10$ for the second game even if we reverse the β-team designation in the design. The reason is that the θth quantile $(\theta \neq 0.5)$ for $S_5 - S_6$ is not the same as the θth quantile $S_6 - S_5$.

5 NFL ratings and rankings

The quantile regression coefficients for deciles 0.1 to 0.9 are presented in Table 3. Also included is the least squares estimate. The intercept for least squares is zero because the home field estimate (3.64) has been subtracted from the outcomes. The quantile intercepts satisfy the symmetry constraint.

Tables 4 and 5 present the ratings and rankings for each team. In terms of the rating values we see that Indianapolis with a score of 12 at the median is 5.4 points better than Chicago which has a 6.6 rating. On the other hand, Indianapolis is 9.3 point favorite in terms of the expected value $(10.8 - 1.5 = 9.3)$. Comparing Indianapolis and Carolina at $\theta = 0.5$ we see that Indianapolis is a 10-point favorite $(12 - 1.6)$ whereas at the expected value it is only about a 5-point favorite.

---

[7] We do not consider the problem of how to best calibrate the weights (how much should early season losses weigh in the determination of end of season ratings?).

Table 3 Regression quantile coefficients

Team	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	E()
Arizona	−1.48	−0.22	−0.89	−3.26	−5.82	−5.58	−7.84	−8.25	−7.18	−5.05
Atlanta	−1.48	2.78	0.16	−2.26	0.18	−0.58	−0.64	3.75	7.64	−1.12
Baltimore	−0.81	−3.37	−5.37	0.07	−2.62	−0.44	−3.64	−5.25	−2.27	−1.82
Buffalo	0.19	−5.22	−5.33	−3.59	−5.82	−5.58	−6.44	−8.25	−6.27	−5.79
Carolina	7.19	5.21	5.16	1.07	1.58	5.38	7.36	8.75	10.59	5.18
Chicago	−5.81	−1.50	1.14	3.74	6.58	−0.58	2.56	5.75	7.64	1.46
Cincinnati	1.52	2.92	3.63	3.07	1.58	5.60	−0.44	−5.25	−6.22	3.84
Cleveland	−10.81	−5.79	−3.40	−3.26	−2.02	−3.40	−5.24	−2.25	1.69	−4.20
Dallas	1.52	3.50	6.65	5.41	3.98	3.47	0.56	−0.25	−3.13	3.05
Denver	13.52	12.50	10.09	8.74	9.38	12.38	13.36	14.75	13.87	10.74
Detroit	−12.81	−11.50	−8.86	−5.59	−2.82	−2.58	−3.44	−3.25	−1.22	−6.63
Green Bay	−3.15	−2.79	−3.35	−3.59	−2.02	−5.62	−0.04	0.75	−8.18	−3.65
Houston	−11.48	−10.22	−10.40	−11.59	−9.42	−8.53	−11.64	−10.25	−12.27	−10.04
Indianapolis	7.85	9.50	9.11	11.41	11.98	16.42	13.96	10.75	4.73	10.79
Jacksonville	3.19	2.07	3.60	3.07	4.58	8.38	8.96	7.75	7.78	4.75
Kansas City	6.85	7.63	5.67	7.41	5.18	7.42	7.36	7.75	10.82	6.95
Miami	1.19	1.50	−1.81	−2.93	−8.22	−4.62	−4.04	−5.25	−3.36	−0.71
Minnesota	−0.48	−5.22	−2.86	−0.26	4.38	−0.49	−3.04	−3.25	−0.41	−3.48
New England	−0.81	2.78	0.65	1.07	0.38	4.38	7.16	10.75	4.73	3.20
New Orleans	−4.15	−7.79	−8.86	−6.93	−8.02	−10.53	−10.24	−9.25	−11.22	−10.60
NY Giants	6.19	6.07	9.60	6.74	6.18	6.42	5.36	4.75	9.73	7.00
NY Jets	−1.15	−4.08	−8.33	−9.26	−9.22	−5.62	−3.84	−5.25	−3.18	−6.34
Oakland	−3.15	−2.08	−2.89	−5.59	−6.22	−5.53	−3.24	−7.25	−2.22	−2.86
Philadelphia	−10.48	−6.93	−2.91	−2.93	−0.42	−4.49	−4.84	−1.25	0.82	−2.41
Pittsburgh	5.85	4.07	7.09	8.07	7.78	4.47	4.76	4.75	2.69	7.83
San Diego	7.19	5.21	7.60	7.74	9.78	8.47	10.96	8.75	14.78	9.88
San Francisco	−7.48	−6.22	−5.42	−8.93	−9.82	−12.49	−15.04	−14.25	−18.27	−11.21
Seattle	14.19	12.21	10.60	9.41	5.78	3.33	6.76	8.75	2.78	9.05
St Louis	−1.81	−4.50	−6.91	−5.26	−4.82	−7.53	−7.24	−4.25	−6.18	−5.19
Tampa Bay	0.85	0.07	0.65	0.41	2.38	−0.67	−0.24	−4.25	−6.31	−0.90
Tennessee	−3.48	−5.37	−9.37	−8.93	−9.82	−8.62	−8.24	−8.25	−10.31	−7.58
Washington	3.52	4.78	5.58	6.74	5.38	7.38	10.16	7.75	7.91	5.88
Constant	−13.31	−9.36	−6.16	−3.31	−0.24	3.31	6.16	9.36	13.31	0.00

Note: Home = 3.64

For some teams the alternative rating criteria lead to similar estimates. There is general agreement that the top three teams are Indianapolis, San Diego and Denver. After the top three teams there are sometimes major differences. For example, Chicago is ranked fifth in the middle of the distribution but much worse in the other parts of the distribution, and its rating based on expected values is 14th. Kansas City on the other hand is ninth best in the middle of the distribution but fourth at the 0.9 quantile. Similarly, Carolina is in the middle of the pack, 14th in the middle, but third in the tail. When Carolina does win, it evidently is by a large margin.

The superbowl winner, Pittsburgh, ranks high under all criteria, but never number one. Perhaps its championship was due to luck, and the more likely champion should have been Indianapolis, San Diego or Denver. But it is also possible that a different calibration would favor Pittsburgh. Recall that the estimates give equal weight to all

Table 4 NFL ratings 2005 regular season

θ	0.5	0.6	0.7	0.8	0.9	E()
s	**−0.2**	**3.3**	**6.2**	**9.4**	**13.3**	
Indianapolis	12.0	13.9	11.5	10.1	6.3	10.8
San Diego	9.8	8.1	9.3	7.0	11.0	9.9
Denver	9.4	10.6	11.7	13.6	13.7	10.7
Pittsburgh	7.8	6.3	5.9	4.4	4.3	7.8
Chicago	6.6	1.6	1.8	2.1	0.9	1.5
NY Giants	6.2	6.6	7.5	5.4	8.0	7.0
Seattle	5.8	6.4	8.7	10.5	8.5	9.0
Washington	5.4	7.1	7.9	6.3	5.7	5.9
Kansas City	5.2	7.4	6.5	7.7	8.8	6.9
Jacksonville	4.6	5.7	6.3	4.9	5.5	4.7
Minnesota	4.4	−0.4	−3.0	−4.2	−0.4	−3.5
Dallas	4.0	4.4	3.6	1.6	−0.8	3.0
Tampa Bay	2.4	−0.1	0.2	−2.1	−2.7	−0.9
Carolina	1.6	3.2	6.3	7.0	8.9	5.2
Cincinnati	1.6	4.3	1.6	−1.2	−2.4	3.8
New England	0.4	2.7	3.9	6.8	2.0	3.2
Atlanta	0.2	−1.4	−0.2	3.3	3.1	−1.1
Philadelphia	−0.4	−3.7	−3.9	−4.1	−4.8	−2.4
Green Bay	−2.0	−4.6	−1.7	−1.0	−5.7	−3.7
Cleveland	−2.0	−3.3	−4.3	−4.0	−4.6	−4.2
Baltimore	−2.6	−0.2	−4.5	−4.3	−1.5	−1.8
Detroit	−2.8	−4.1	−6.2	−7.4	−7.0	−6.6
St Louis	−4.8	−6.4	−7.1	−4.4	−4.0	−5.2
Arizona	−5.8	−4.4	−4.4	−4.2	−4.3	−5.0
Buffalo	−5.8	−4.6	−5.9	−6.7	−3.0	−5.8
Oakland	−6.2	−5.6	−3.1	−4.7	−2.7	−2.9
New Orleans	−8.0	−8.7	−9.6	−8.5	−7.7	−10.6
Miami	−8.2	−3.8	−2.9	−1.9	−1.1	−0.7
NY Jets	−9.2	−7.4	−6.1	−4.7	−2.2	−6.3
Houston	−9.4	−10.1	−11.0	−10.2	−11.9	−10.0
San Francisco	−9.8	−10.7	−10.2	−10.2	−12.9	−11.2
Tennessee	−9.8	−8.8	−8.8	−6.8	−6.9	−7.6

Table 5 NFL rankings 2005 regular season

θ	0.5	0.6	0.7	0.8	0.9	E()
s	−0.2	3.3	6.2	9.4	13.3	
Indianapolis	1	1	2	3	7	1
San Diego	2	3	3	5	2	3
Denver	3	2	1	1	1	2
Pittsburgh	4	8	10	11	10	5
Chicago	5	14	13	13	13	14
NY Giants	6	6	6	9	6	6
Seattle	7	7	4	2	5	4
Washington	8	5	5	8	8	8
Kansas City	9	4	7	4	4	7
Jacksonville	10	9	8	10	9	10
Minnesota	11	17	19	21	14	21
Dallas	12	10	12	14	15	13
Tampa Bay	13	15	15	18	21	16
Carolina	14	12	9	6	3	9
Cincinnati	14	11	14	16	19	11
New England	16	13	11	7	12	12
Atlanta	17	18	16	12	11	17
Philadelphia	18	20	21	20	26	19
Green Bay	19	25	17	15	27	22
Cleveland	20	19	22	19	25	23
Baltimore	21	16	24	23	17	18
Detroit	22	22	27	29	29	28
St Louis	23	27	28	24	23	25
Arizona	24	23	23	22	24	24
Buffalo	24	24	25	27	22	26
Oakland	26	26	20	25	20	20
New Orleans	27	29	30	30	30	31
Miami	28	21	18	17	16	15
NY Jets	29	28	26	25	18	27
Houston	30	31	32	31	31	30
San Francisco	31	32	31	31	32	32
Tennessee	31	30	29	28	28	29

outcomes. Games played in the first week of the season are just as important as games in the last week. To win playoff games and eventually the Super Bowl, it is plausible that late-season games are a better indicator of team strength so that greater weight on late-season games would give a better ranking of teams at the end of the season. Calibrating weights is a topic for further study.[8]

[8] Another topic for future research is how to combine the quantile estimates into a single ranking. For example, it might be reasonable to base rankings on the outcome of a round robin. The NFL schedule is not a round robin, but with the quantile estimates we can estimate and then derive a ranking based on the expected number of wins for each team in a round robin tournament.

6 Summary

We have shown how team ratings and rankings depend on where in the distribution the rating problem is formulated. Teams rated best on one criterion can be lower on another. We have shown how the quantile regression model can be used to formulate the rating question and estimate the ratings based on game outcomes.

Given the sometimes conflicting answers to the rating question, it is tempting to ask who *really* is best. How can the ratings be aggregated into an answer that declares which truly is the top team? While a single rating answer does emerge from the standard approach, it is an artifact of the location-shift model. This contrasts with the quantile regression approach that allows the data to determine the ratings and rankings. In the end, the answer to who is best, of where to place your bets depends on where in the distribution outcomes are compared.

Acknowledgement

An earlier version of this paper was presented at the workshop on Quantile Regression, LMS Method and Robust Statistics in the 21st Century, International Centre for Mathematical Sciences, Edinburgh, June 2006. Thanks to the conference participants as well as an anonymous reviewer for very helpful comments.

References

Bassett GW Jr (1997) Robust sports ratings based on least absolute values. *American Statistician*, **51**, 99–105.

Bassett GW (1981) Point spreads vs. odds. *Journal of Political Economy*, **80**, 752–68.

David HA (1988) The method of paired comparisons, second edition, revised. London: Oxford University Press.

Frey J (2005) A ranking method based on minimizing the number of in-sample errors. *The American Statistician*, **59**, 207–16.

Harville D (1977) The use of linear-model methodology to rate high school or college football teams. *Journal of the American Statistical Association*, **72**, 278–89.

Koenker R (1984) A note on L-estimators for linear models. *Statistics and Probability Letters*, **2**, 323–25.

Koenker R (2005) *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker R and Hallock K (2001) Quantile regression. *Journal of Economic Perspectives*, **15**, 143–56.

Koenker R and Bassett G (1978) Regression quantiles. *Econometrica*, **46**, 33–50.

Massey K (2006) http://www.masseyratings.com (accessed 21 march 2006).

Stern H (1995) Who's number one in college football?... And how might we decide? *Chance*, **8**(3), 7–14.

Stern H (2004) Statistics and the college football championship. *The American Statistician*, **58**, 179–85.