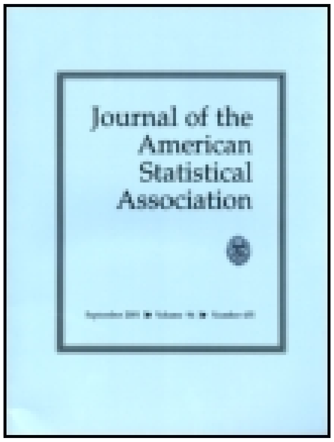

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Predictions for National Football League Games via Linear-Model Methodology

David Harville [a]

[a] Department of Statistics , Iowa State University , Ames , IA , 50011 , USA

Published online: 12 Mar 2012.

Predictions for National Football League Games Via Linear-Model Methodology

DAVID HARVILLE*

Results on mixed linear models were used to develop a procedure for predicting the outcomes of National Football League games. The predictions are based on the differences in score from past games. The underlying model for each difference in score takes into account the home-field advantage and the difference in the yearly characteristic performance levels of the two teams. Each team's yearly characteristic performance levels are assumed to follow a first-order autoregressive process. The predictions for 1,320 games played between 1971 and 1977 had an average absolute error of 10.68, compared with 10.49 for bookmaker predictions.

KEY WORDS: Football predictions; Mixed linear models; Variance components; Maximum likelihood; Football ratings.

1. INTRODUCTION

Suppose that we wish to predict the future price of a common stock or to address some other complex real-life prediction problem. What is the most useful role for statistics?

One approach is to use the available information rather informally, relying primarily on intuition and on past experience and employing no statistical methods or only relatively simple statistical methods. A second approach is to rely exclusively on some sophisticated statistical algorithm to produce the predictions from the relevant data. In the present article, these two approaches are compared in the context of predicting the outcomes of National Football League (NFL) games.

The statistical algorithm to be used is set forth in Section 2. It is closely related to an algorithm devised by Harville (1977b) for rating high school or college football teams.

The essentially nonstatistical predictions that are to be compared with the statistical predictions are those given by the betting line. The betting line gives the favored team for each game and the point spread, that is, the number of points by which the favorite is expected to win.

If a gambler bets on the favorite (underdog), he wins (loses) his bet when the favorite wins the game by more than the point spread, but he loses (wins) his bet when the favorite either loses or ties the game or wins the game by less than the point spread. On a \$10 bet, the gambler pays the bookmaker an additional dollar (for a total of \$11) when he loses his bet and receives a net of \$10 when he wins. If the favorite wins the game by exactly the point spread, the bet is in effect cancelled (Merchant 1973). To break even, the gambler must win 52.4 percent of those bets that result in either a win or a loss (assuming that the bets are for equal amounts).

Merchant described the way in which the betting line is established. A prominent bookmaker devises an initial line, which is known as the outlaw line, the early line, or the service line. This line is such that, in his informed opinion, the probability of winning a bet on the favorite equals the probability of winning a bet on the underdog.

A select group of knowledgeable professional gamblers are allowed to place bets (in limited amounts) on the basis of the outlaw line. A series of small adjustments is made in the outlaw line until an approximately equal amount of the professionals' money is being attracted on either side. The betting line that results from this process is the official opening line, which becomes available on Tuesday for public betting.

Bets can be placed until the game is played, which is generally on Sunday but can be as early as Thursday or as late as Monday. If at any point during the betting period the bookmaker feels that there is too big a discrepancy between the amount being bet on the favorite and the amount being bet on the underdog, he may make a further adjustment in the line.

The nonstatistical predictions used in the present study are those given by the official opening betting line. These predictions can be viewed as the consensus opinion of knowledgeable professional gamblers.

The statistical algorithm that is set forth in Section 2 can be used to rate the various NFL teams as well as to make predictions. While the prediction and rating problems are closely related, there are also some important differences, which are discussed in Section 5.

2. STATISTICAL PREDICTIONS

Each year's NFL schedule consists of three parts: preseason or exhibition games, regular-season games, and postseason or playoff games. The statistical algorithm presented in Sections 2.2 through 2.4 translates scores from regular-season and playoff games that have already been played into predictions for regular-season and play-

* David Harville is Professor, Department of Statistics, Iowa State University, Ames, IA 50011. This article is based on an invited paper (Harville 1978) presented at the 138th Annual Meeting of the American Statistical Association, San Diego, CA.

September 1980, Volume 75, Number 371
Applications Section

off games to be played in the future. The model that serves as the basis for this algorithm is described in Section 2.1.

The scores of exhibition games were not used in making the statistical predictions, and predictions were not attempted for future exhibition games. The rationale was that these games are hard to incorporate into the model and thus into the algorithm and that they have very little predictive value anyhow. Merchant (1973) argues that, in making predictions for regular-season games, it is best to forget about exhibition games.

2.1 Underlying Model

Suppose that the scores to be used in making the predictions date back to the Year F. Ultimately, F is to be chosen so that the interlude between the beginning of Year F and the first date for which predictions are required is long enough that the effect of including earlier scores is negligible. For each year, number the regular-season and playoff games 1, 2, 3, ... in chronological order.

Number the NFL teams 1, 2, 3, ... New teams are formed by the NFL from time to time. If Team i is added after Year F, let $F(i)$ represent the year of addition. Otherwise, put $F(i) = F$. The home team and the visiting team for the kth game in Year j are to be denoted by $h(j, k)$ and $v(j, k)$, respectively. (If the game were played on a neutral field, $h(j, k)$ is taken arbitrarily to be one of the two participating teams, and $v(j, k)$ is taken to be the other.)

Let S_{jk} equal the home team's score minus the visiting team's score for the kth game in Year j. The prediction algorithm presented in Sections 2.2 through 2.4 depends only on the scores and depends on the scores only through the S_{jk}'s.

Our model for the S_{jk}'s involves conceptual quantities H and T_{im} ($i = 1, 2, \ldots; m = F(i), F(i) + 1, \ldots$). The quantity H is an unknown parameter that represents the home-field advantage (in points) that accrues to a team from playing on its own field rather than a neutral field. The quantity T_{im} is a random effect that can be interpreted as the characteristic performance level (in points) of Team i in Year m relative to that of an "average" team in Year m.

The model equation for S_{jk} is

$$S_{jk} = T_{h(j,k),j} - T_{v(j,k),j} + R_{jk}$$

if the game is played on a neutral field, or

$$S_{jk} = H + T_{h(j,k),j} - T_{v(j,k),j} + R_{jk}$$

if it is not. Here, R_{jk} is a random residual effect. Assume that $E(R_{jk}) = 0$, that $\text{var}(R_{jk}) = \sigma_R^2$ where σ_R^2 is an unknown strictly positive parameter, and that the R_{jk}'s are uncorrelated with each other and with the T_{im}'s.

Suppose that $\text{cov}(T_{im}, T_{i'm'}) = 0$ if $i' \neq i$; that is, that the yearly characteristic performance levels of any given team are uncorrelated with those of any other team. The yearly characteristic performance levels of Team i are assumed to follow a first-order autoregressive process

$$T_{i,m+1} = \rho T_{im} + U_{im} \quad (m = F(i), F(i) + 1, \ldots) ,$$

where $U_{i,F(i)}, U_{i,F(i)+1}, \ldots$ are random variables that have zero means and common unknown variance σ_U^2 and that are uncorrelated with each other and with $T_{i,F(i)}$, and where ρ is an unknown parameter satisfying $0 \leq \rho < 1$.

It remains to specify assumptions, for each i, about $E[T_{i,F(i)}]$ and $\text{var}[T_{i,F(i)}]$, that is, about the mean and variance of the first yearly characteristic performance level for Team i. The sensitivity of the prediction procedure to these specifications depends on the proximity (in time) of the predicted games to the beginning of Year $F(i)$ and on whether the predicted games involve Team i. For i such that $F(i) = F$, that is, for teams that date back to Year F, the prediction procedure will be relatively insensitive to these specifications, provided that F is sufficiently small, that is, provided the formation of the data base was started sufficiently in advance of the first date for which predictions are required.

Put $\sigma_T^2 = \sigma_U^2/(1 - \rho^2)$, and for convenience assume that

$$E[T_{i,F(i)}] = 0 \quad \text{and} \quad \text{var}[T_{i,F(i)}] = \sigma_T^2 \tag{2.1}$$

for i such that $F(i) = F$. Then, for any given year, the yearly characteristic performance levels of those teams that date back to Year F have zero means and common variance σ_T^2, as would be the case if they were regarded as a random sample from an infinite population having mean zero and variance σ_T^2. Moreover, for any such team, the correlation between its characteristic performance levels for any two years m and m' is $\rho^{|m'-m|}$, which is a decreasing function of elapsed time.

For i such that $F(i) > F$, that is, for teams that came into being after Year F, it is assumed that

$$E[T_{i,F(i)}] = \mu_{F(i)} \quad \text{and} \quad \text{var}[T_{i,F(i)}] = \tau_{F(i)}^2 ,$$

where $\mu_{F(i)}$ and $\tau_{F(i)}^2$ are quantities that are to be supplied by the user of the prediction procedure. The quantities $\mu_{F(i)}$ and $\tau_{F(i)}^2$ can be regarded as the mean and variance of a common prior distribution for the initial yearly characteristic performance levels of expansion teams. Information on the performance of expansion teams in their first year that predates Year $F(i)$ can be used in deciding on values for $\mu_{F(i)}$ and $\tau_{F(i)}^2$.

The model for the S_{jk}'s is similar to that applied to high school and college football data by Harville (1977b). One distinguishing feature is the provision for data from more than one year.

2.2 Preliminaries

We consider the problem of predicting S_{JK} from S_{F1}, $S_{F2}, \ldots, S_{LG}$, where either $J = L$ and $K > G$ or $J > L$; that is, the problem of predicting the winner and the margin of victory for a future game based on the information accumulated as of Game G in Year L. This problem is closely related to that of estimating or predicting H and T_{im} ($i = 1, 2, \ldots$; $m = F(i), F(i) + 1, \ldots$).

Take $\lambda = \sigma_T^2/\sigma_R^2$. If λ and ρ were given, H would have a unique minimum-variance linear unbiased (Aitken) estimator, which we denote by $\tilde{H}(\lambda, \rho)$. Define $\tilde{T}_{im}(\lambda, \rho, H)$ by

$$\tilde{T}_{im}(\lambda, \rho, H) = E(T_{im} | S_{F1}, S_{F2}, \ldots, S_{LG}) \ , \quad (2.2)$$

where the conditional expectation is taken under the assumption that T_{im} and $S_{F1}, S_{F2}, \ldots, S_{LG}$ are jointly normal or, equivalently, is taken to be Hartigan's (1969) linear expectation. Put

$$\tilde{T}_{im}(\lambda, \rho) = \tilde{T}_{im}(\lambda, \rho, \tilde{H}(\lambda, \rho)) \ . \quad (2.3)$$

The quantity $\tilde{T}_{im}(\lambda, \rho)$ is the best linear unbiased predictor (BLUP) of T_{im} (in the sense described by Harville 1976) for the case in which λ and ρ are given; and, for that same case, the quantity $\tilde{S}_{JK}(\lambda, \rho)$, defined as follows is the BLUP of S_{jk}:

$$\tilde{S}_{JK}(\lambda, \rho) = \tilde{T}_{h(J,K),J}(\lambda, \rho) - \tilde{T}_{v(J,K),J}(\lambda, \rho) \quad (2.4)$$

if the JKth game is played on a neutral field; or

$$\tilde{S}_{JK}(\lambda, \rho) = \tilde{H}(\lambda, \rho) + \tilde{T}_{h(J,K),J}(\lambda, \rho) - \tilde{T}_{v(J,K),J}(\lambda, \rho) \quad (2.5)$$

if it is not.

Let $M_{JK}(\sigma_R^2, \lambda, \rho)$ denote the mean squared difference between $\tilde{S}_{JK}(\lambda, \rho)$ and S_{JK}, that is,

$$M_{JK}(\sigma_R^2, \lambda, \rho) = E[\tilde{S}_{JK}(\lambda, \rho) - S_{JK}]^2 \ . \quad (2.6)$$

Specific representations for $\tilde{H}(\lambda, \rho)$, $\tilde{T}_{im}(\lambda, \rho)$, and $M_{JK}(\sigma_R^2, \lambda, \rho)$ can be obtained as special cases of representations given, for example, by Harville (1976).

2.3 Estimation of Model Parameters

In practice, λ and ρ (and σ_R^2) are not given and must be estimated. One approach to the estimation of these parameters is to use Patterson and Thompson's (1971) restricted maximum likelihood procedure. (See, e.g., Harville's (1977a) review article for a general description of this procedure.)

Suppose that $S_{F1}, S_{F2}, \ldots, S_{LG}$ constitute the data available for estimating λ, ρ, and σ_R^2. For purposes of estimating these parameters, we assume that the data are jointly normal, and we take $E[T_{i,F(i)}]$ and $\text{var}[T_{i,F(i)}]$ to be of the form (2.1) for *all* i; however, we eliminate from the data set any datum S_{jk} for which $h(j, k)$ or $v(j, k)$ corresponds to an expansion team formed within I years of Year j (where I is to be specified by the user). Extending the assumption (2.1) to all i simplifies the estimation procedure, while the exclusion of games involving expansion teams in their early years desensitizes the procedure to the effects of this assumption.

We write $\tilde{H}$ and $\tilde{T}_{im}$ for the quantities $\tilde{H}(\lambda, \rho)$ and $\tilde{T}_{im}(\lambda, \rho)$ defined in Section 2.2 (with allowances for the deletions in the data set and the change in assumptions). Let X_{jk} equal 0 or 1 depending on whether or not the jkth game is played on a neutral field.

The likelihood equations for the restricted maximum likelihood procedure can be put into the form

$$Q_i - E(Q_i) = 0 \quad (i = 1, 2, 3) \ , \quad (2.7)$$

with

$$Q_1 = (\tfrac{1}{2})[\sigma_R^2\lambda(1-\rho^2)]^{-1} \sum_i [(1-\rho^2)\tilde{T}_{i,F(i)} + \sum_{k=F(i)}^{L-1} (\tilde{T}_{i,k+1} - \rho\tilde{T}_{ik})^2] \ , \quad (2.8)$$

$$Q_2 = (\tfrac{1}{2})[\sigma_R^2\lambda(1-\rho^2)]^{-1} \sum_i \sum_{k=F(i)}^{L-1} (\tilde{T}_{ik} - \rho\tilde{T}_{i,k+1})(\tilde{T}_{i,k+1} - \rho\tilde{T}_{ik}) \ , \quad (2.9)$$

and

$$Q_3 = \sum_{j=F}^{L} \sum_k [S_{jk} - X_{jk}\tilde{H} - \tilde{T}_{h(j,k),j} + \tilde{T}_{v(j,k),j}]^2 \ . \quad (2.10)$$

It can be shown that

$$E[S_{jk} - X_{jk}\tilde{H} - \tilde{T}_{h(j,k),j} + \tilde{T}_{v(j,k),j}]^2 = \sigma_R^2 - (X_{jk}, 1, -1) \cdot \text{var}[\tilde{H} - H, \tilde{T}_{h(j,k),j} - T_{h(j,k),j}, \tilde{T}_{v(j,k),j} - T_{v(j,k),j}] \cdot (X_{jk}, 1, -1)' \quad (2.11)$$

and that

$$E(\tilde{T}_{im}) = E(T_{im}) = 0 \quad (2.12)$$

and

$$\text{cov}(\tilde{T}_{im}, \tilde{T}_{im'}) = \rho^{|m'-m|}\sigma_T^2 - \text{cov}(\tilde{T}_{im} - T_{im}, \tilde{T}_{im'} - T_{im'}) \ . \quad (2.13)$$

Equations (2.7) can be solved numerically by the same iterative numerical algorithm used by Harville (1977b, p. 288). This procedure calls for the repeated evaluation of the Q_i's and their expectations for various trial values of λ and ρ. Making use of (2.11), (2.12), and (2.13) reduces the problem of evaluating the Q_i's and their expectations for particular values of λ and ρ to the problem of evaluating $\tilde{H}$ and the $\tilde{T}_{im}$'s and various elements of their dispersion matrix. Kalman filtering and smoothing algorithms (suitably modified for mixed models, as described by Harville 1979) can be used for maximum efficiency in carrying out the computations associated with the latter problem.

The amount of computation required to evaluate $\tilde{H}$ and the $\tilde{T}_{im}$'s and the relevant elements of their dispersion matrix, for fixed values of λ and ρ, may not be feasible if data from a large number of years are being used. The procedure for estimating λ, ρ, and σ_R^2 can be modified in these instances by, for example, basing the "estimates" $\tilde{T}_{i,k+1}$ and $\tilde{T}_{ik}$ in the term $(\tilde{T}_{i,k+1} - \rho\tilde{T}_{ik})^2$ of (2.8) on only those data accumulated through Year $k + Y$, for some Y, rather than on all the data. Such modifications can significantly reduce the amount of storage and computation required to evaluate the Q_i's and their expectations. (The modifications reduce the amount of smoothing that must be carried out in the Kalman algorithm.)

This modified estimation procedure can be viewed as a particular implementation of the approximate restricted maximum likelihood approach outlined by Harville (1977a, Sec. 7). This approach seems to have produced a reasonable procedure even though it is based on the assumption of a distributional form (multivariate normal) that differs considerably from the actual distributional form of the S_{jk}'s.

2.4 Prediction Algorithm

Let $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$ represent estimates of λ, ρ, and σ_R^2, respectively. In particular, we can take $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$ to be the estimates described in Section 2.3.

Let

$$\hat{H} = \tilde{H}(\hat{\lambda}, \hat{\rho}) \quad \text{and} \quad \hat{T}_{im} = \tilde{T}_{im}(\hat{\lambda}, \hat{\rho}) \quad (i = 1, 2, \ldots; m = F(i), F(i) + 1, \ldots) .$$

The quantity $\hat{H}$ gives an estimate of H, and $\hat{T}_{im}$ gives an estimate or prediction of T_{im}.

It can be shown that

$$\hat{H} = N^{-1}(G - \sum_i \sum_{m=F(i)}^{L} D_{im}\hat{T}_{im}) , \qquad (2.14)$$

where N equals the total number of games played minus the number of games played on neutral fields, G equals the grand total of all points scored by home teams minus the grand total for visiting teams, and D_{im} equals the number of games played in Year m by Team i on its home field minus the number played on its opponents' fields. The NFL schedule is such that, if it were not for playoff games and for games not yet played in Year L, all of the D_{im}'s would equal zero, and $\hat{H}$ would coincide with the ordinary average $N^{-1}G$.

It can also be shown that, if $F(i) = L$; that is, if only one season of data or a partial season of data is available on Team i, then

$$\hat{T}_{iL} = E(T_{iL}) + (N_{iL} + \hat{\gamma})^{-1} \cdot [G_{iL} - N_{iL}E(T_{iL}) - D_{iL}\hat{H} + \sum_j \hat{T}_{r(j),L}] , \qquad (2.15)$$

where N_{iL} equals the number of games played (in Year L) by Team i, G_{iL} equals the total points scored (in Year L) by Team i minus the total scored against it by its opponents, $r(j)$ equals Team i's opponent in its jth game (of Year L), and

$\hat{\gamma}$ equals $\hat{\lambda}^{-1}$ if $F(i) = F$ and equals $\hat{\sigma}_R^2/\tau_{F(i)}^2$ if $F(i) > F$.

Thus, if $F(i) = L$, the estimator $\hat{T}_{iL}$ is seen to be a "shrinker"; that is, instead of the "corrected total" for the ith team being divided by N_{iL}, it is divided by $N_{iL} + \hat{\gamma}$. If $F(i) < L$, that is, if more than one season of data is available on Team i, the form of the estimator is similar.

The prediction for the outcome S_{JK} of a future game is taken to be

$$\hat{S}_{JK} = \tilde{S}_{JK}(\hat{\lambda}, \hat{\rho}) = X_{JK}\hat{H} + \hat{T}_{h(J,K),J} - \hat{T}_{v(J,K),J} .$$

An estimate of the mean squared error of this prediction is given by $\hat{M}_{JK} = M_{JK}(\hat{\sigma}_R^2, \hat{\lambda}, \hat{\rho})$. The estimate $\hat{M}_{JK}$ underestimates the mean squared error to an extent that depends on the precision of the estimates $\hat{\lambda}$ and $\hat{\rho}$.

The quantities $\tilde{S}_{JK}(\lambda, \rho)$ and $M_{JK}(\sigma_R^2, \lambda, \rho)$ can be interpreted as the mean and the variance of a posterior distribution for S_{JK} (Harville 1976, Sec. 4). Depending on the precision of the estimates $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$, it may be reasonable to interpret $\hat{S}_{JK}$ and $\hat{M}_{JK}$ in much the same way.

Take B_{JK} to be a constant that represents the difference in score given by the betting line for Game K of Year J. Relevant posterior probabilities for gambling purposes are $\Pr(S_{JK} < B_{JK})$ and $\Pr(S_{JK} > B_{JK})$. These posterior probabilities can be obtained from the posterior probabilities $\Pr(S_{JK} = s)$ $(s = \ldots, -2, -1, 0, 1, 2, \ldots)$, which we approximate by $\Pr(s - .5 < S_{JK}^* \leq s + .5)$, where S_{JK}^* is a normal random variable with mean $\hat{S}_{JK}$ and variance $\hat{M}_{JK}$.

Due to the "lumpiness" of the actual distribution of the differences in score (as described, for college football, by Mosteller 1970) these approximations to the posterior probabilities may be somewhat crude; however, it is not clear how to improve on them by other than ad hoc procedures. Rosner (1976) took a somewhat different approach to the prediction problem in an attempt to accommodate this feature of the distribution.

Our prediction algorithm can be viewed as consisting of two stages. In the first stage, λ, ρ, and σ_R^2 are estimated. Then, in the second stage, $\hat{H}$ and $\hat{T}_{iL}$ $(i = 1, 2, \ldots)$ and their estimated dispersion matrix are computed. By making use of the Kalman prediction algorithm (as described by Harville 1979), the output of the second stage can easily be converted into a prediction $\hat{S}_{JK}$ and an estimated mean squared prediction error $\hat{M}_{JK}$ for any future game.

The second-stage computations, as well as the first-stage computations, can be facilitated by use of the Kalman filtering algorithm. This is especially true in instances where the second-stage computations were previously carried out based on S_{jk}'s available earlier and where λ and ρ have not been reestimated.

3. EMPIRICAL EVALUATION OF PREDICTIONS

The statistical algorithm described in Section 2 was used to make predictions for actual NFL games. These predictions were compared for accuracy with those given by the betting line. The games for which the comparisons were made were 1,320 regular-season and playoff games played between 1971 and 1977, inclusive.

The betting line for each game was taken to be the opening line. The primary source for the opening line was the *San Francisco Chronicle*, which, in its Wednesday

1. Parameter Estimates Over Each of Seven Time Periods

Last Year in Period	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\sigma}_R^2$	$\hat{H}$
1970	.29	.79	185	—
1971	.26	.83	181	2.19 ± .50
1972	.27	.81	180	2.03 ± .44
1973	.28	.82	182	2.27 ± .40
1974	.26	.78	175	2.31 ± .37
1975	.27	.80	175	2.18 ± .34
1976	.25	.79	171	2.32 ± .32
1977	—	—	—	2.42 ± .30

editions, ordinarily reported the opening line listed in *Harrah's Tahoe Racebook*. There were 31 games played between 1971 and 1977 for which no line could be found and for which no comparisons were made.

The statistical prediction for each of the 1,320 games was based on the outcomes of all NFL regular-season and playoff games played from the beginning of the 1968 season through the week preceding the game to be predicted. The estimates $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$ used in making the predictions were based on the same data but were recomputed yearly rather than weekly. (In estimating λ, ρ, and σ_R^2 at the ends of Years 1970–1975, I was taken to be zero, so that all accumulated games were used. However, in estimating these parameters at the end of 1976, games played during 1968 were excluded as were games involving the expansion teams, Seattle and Tampa Bay, that began play in 1976.)

Games that were tied at the end of regulation play and decided in an overtime period were counted as ties when used in estimating λ, ρ, and σ_R^2 and in making predictions. The values assigned to $\mu_{F(i)}$ and $\tau_{F(i)}^2$ for prediction purposes were -11.8 and 17.0, respectively.

The values obtained for $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$ at the end of each year are listed in Table 1. The values of the estimate $\hat{H}$ of the home-field advantage and the estimated standard error of $\hat{H}$ as of the end of each year (based on the values $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\sigma}_R^2$ obtained at the end of the previous year) are also given.

Some decline in the variability of the outcomes of the games (both "among teams" and "within teams") appears to have taken place beginning in about 1974. The estimates of the among-teams variance σ_T^2 and the within-teams variance σ_R^2 are both much smaller than those obtained by Harville (1977b) for Division-I college football (42 and 171 vs. 104 and 214, respectively). The estimate of the home-field advantage is also smaller than for college football (2.42 vs. 3.35). Merchant (1973) conjectured that, in professional football, the home-field advantage is becoming a thing of the past. The results given in Table 1 seem to indicate otherwise.

Table 2 provides comparisons, broken down on a year-by-year basis, between the accuracy of the statistical predictions and the accuracy of the predictions given by the betting line. Three criteria were used to assess accuracy: the frequency with which the predicted winners actually won, the average of the absolute values of the prediction errors, and the average of the squares of the prediction errors, where prediction error is defined to be the actual (signed) difference in score between the home team and the visiting team minus the predicted difference. With regard to the first criterion, a predicted tie was counted as a success or half a success depending on whether the actual outcome was a tie, and, if a tie occurred but was not predicted, credit for half a success was given.

The statistical predictions are seen to be somewhat less accurate on the average than the predictions given by the betting line. Comparisons with Harville's (1977b) results indicate that both types of predictions tend to be more accurate for professional football than for college football. The average absolute difference between the statistical predictions and the predictions given by the betting line was determined to be 2.48.

Table 3 gives the accuracy, as measured by average absolute error, of the two types of predictions for each of the 14 weeks of the regular season and for the playoff games. Both types of predictions were more accurate over a midseason period, extending approximately from Week 6 to 13, than at the beginning or end of the season. Also, the accuracy of the statistical predictions compared more favorably with that of the betting line during midseason than during the rest of the season. Specifically, the average absolute prediction error for Weeks 6 through 13 was 10.37 for the statistical predictions and 10.35 for the betting line (vs. the overall figures of 10.68 and 10.49, respectively).

2. Accuracy of the Statistical Procedure Versus That of the Betting Line

Year(s)	Number of Games	Percentage of Winners: Statistical Procedure	Percentage of Winners: Betting Line	Average Absolute Error: Statistical Procedure	Average Absolute Error: Betting Line	Average Squared Error: Statistical Procedure	Average Squared Error: Betting Line
1971	164	66.2	68.6	10.28	10.61	172.1	181.0
1972	189	66.4	71.4	11.35	10.94	201.8	192.6
1973	187	74.6	75.7	11.89	11.36	228.2	205.6
1974	187	65.5	68.5	10.07	10.16	159.8	161.6
1975	187	73.8	76.2	10.76	10.38	201.0	183.6
1976	203	72.7	72.9	10.79	10.60	185.9	182.6
1977	203	72.4	70.9	9.67	9.46	166.1	168.0
All	1,320	70.3	72.1	10.68	10.49	187.8	182.0

3. *Week-by-Week Breakdown for Prediction Accuracy*

Week(s)	Number of Games	Average Absolute Error: Statistical Procedure	Average Absolute Error: Betting Line
1	92	11.55	11.20
2	80	10.46	9.79
3	93	11.31	11.31
4	92	10.97	10.39
5	92	11.02	10.52
6	93	10.41	10.27
7	92	10.99	10.72
8	93	10.48	10.78
9	93	8.95	8.94
10	93	9.82	9.95
11	91	10.34	10.26
12	93	10.02	10.12
13	93	11.96	11.75
14	84	11.73	11.18
Playoffs	46	9.96	9.73
All	1,320	10.68	10.49

It is not surprising that the statistical predictions are more accurate, relative to the betting line, during mid-season than during earlier and later periods. The statistical predictions are based only on differences in score from previous games. A great deal of additional information is undoubtedly used by those whose opinions are reflected in the betting line. The importance of taking this additional information into account depends on the extent to which it is already reflected in available scores. Early and late in the season, the additional information is less redundant than at mid-season. At the beginning of the season, it may be helpful to supplement the information on past scores with information on roster changes, injuries, exhibition-game results, and so on. During the last week or two of the regular season, it may be important to take into account which teams are still in contention for playoff berths.

The statistical predictions were somewhat more similar to the predictions given by the betting line during mid-season than during earlier and later periods. For Weeks 6 through 13, the average absolute difference between the two types of predictions was found to be 2.27 (vs. the overall figure of 2.48).

There remains the question of whether the statistical predictions could serve as the basis for a successful betting scheme. Suppose that we are considering betting on a future game, say, Game K of Year J.

If $\hat{S}_{JK} > B_{JK}$, that is, if the statistical prediction indicates that the chances of the home team are better than those specified by the betting line, then we might wish to place a bet on the home team. The final decision on whether to make the bet could be based on the approximation (discussed in Section 2.4) to the ratio

$$\Pr(S_{JK} > B_{JK}) / [\Pr(S_{JK} > B_{JK}) + \Pr(S_{JK} < B_{JK})] , \quad (3.1)$$

that is, on the approximation to the conditional probability (as defined in Section 2.4) that the home team will win by more (or lose by less) than the betting line would indicate given that the game does not end in a tie relative to the betting line. The bet would be made if the approximate value of this conditional probability were sufficiently greater than .5.

If $\hat{S}_{JK} < B_{JK}$, then depending on whether the approximation to the ratio (3.1) were sufficiently smaller than .5, we would bet on the visiting team. The actual success of such a betting scheme would depend on the frequency with which bets meeting our criteria arise and on the relative frequency of winning bets among those bets that do qualify.

In Table 4a, the predictions for the 1,320 games are divided into six categories depending on the (approximate) conditional probability that the team favored by

4. *Theoretical Versus Observed Frequency of Success for Statistical Predictions Relative to the Betting Line*

Probability Interval	Number of Games and Number of Ties	Average Probability	Observed Relative Frequency	Number of Games and Number of Ties (Cumulative)	Average Probability (Cumulative)	Observed Relative Frequency (Cumulative)
			a. All weeks			
[.50, .55)	566 (16)	.525	.525	1320 (48)	.570	.528
[.55, .60)	429 (17)	.574	.534	754 (32)	.604	.530
[.60, .65)	221 (12)	.621	.483	325 (15)	.643	.526
[.65, .70)	78 (3)	.671	.627	104 (3)	.688	.614
[.70, .75)	18 (0)	.718	.556	26 (0)	.736	.577
>.75	8 (0)	.778	.625	8 (0)	.778	.625
			b. Weeks 6–13			
[.50, .55)	337 (7)	.525	.503	741 (26)	.564	.541
[.55, .60)	248 (11)	.573	.570	404 (19)	.598	.574
[.60, .65)	112 (6)	.620	.528	156 (8)	.638	.581
[.65, .70)	39 (2)	.672	.730	44 (2)	.682	.714
[.70, .75)	3 (0)	.728	.333	5 (0)	.758	.600
>.75	2 (0)	.805	1.000	2 (0)	.805	1.000

the statistical algorithm relative to the betting line will win by more (or lose by less) than indicated by the betting line (given that the game does not end in a tie relative to the betting line). For each category, the table gives the total number of games or predictions, the number of games that actually ended in a tie relative to the betting line, the average of the conditional probabilities, and the observed frequency with which the teams favored by the statistical algorithm relative to the betting line actually won by more (or lost by less) than predicted by the line (excluding games that actually ended in a tie relative to the line). Cumulative figures, starting with the category corresponding to the highest conditional probabilities, are also given. Table 4b gives the same information for the 741 games of Weeks 6 through 13.

The motivation for the proposed betting scheme is a suspicion that the relative frequency with which the teams favored by the statistical algorithm relative to the line actually beat the line might be a strictly increasing (and possibly approximately linear) function of the "theoretical" frequency (approximate conditional probability), having a value of .50 at a theoretical relative frequency of .50. The results given in Table 4 tend to support this suspicion and to indicate that the rate of increase is greater for the midseason than for the entire season.

Fitting a linear function to overall observed relative frequencies by iterative weighted least squares produced the following equations:

$$\text{relative frequency} = .50 + .285\,(\text{theoretical frequency} - .50) \; .$$

The fitted equation for Weeks 6 through 13 was:

$$\text{relative frequency} = .50 + .655\,(\text{theoretical frequency} - .50) \; .$$

The addition of quadratic and cubic terms to the equations resulted in only negligible improvements in fit.

The proposed betting scheme would generally have shown a profit during the 1971–1977 period. The rate of profit would have depended on whether betting had been restricted to midseason games and on what theoretical frequency had been used as the cutoff point in deciding whether to place a bet. Even if bets (of equal size) had been placed on every one of the 1,320 games, some profit would have been realized (since the overall observed relative frequency was .528 vs. the break-even point of .524).

4. DISCUSSION

4.1 Modification of the Prediction Algorithm

One approach to improving the accuracy of the statistical predictions would be to modify the underlying model. In particular, instead of assuming that the residual effects are uncorrelated, we could, following Harville (1977b), assume that

$$R_{ik} = C_{h(j,k),j,w(j,k)} - C_{v(j,k),j,w(j,k)} + F_{jk} \; , \qquad (4.1)$$

where, taking the weeks of each season to be numbered $1, 2, 3, \ldots, w(j, k) = m$ if Game k of Year j were played during Week m. The quantities C_{imn} and F_{jk} represent random variables such that $E(C_{imn}) = E(F_{jk}) = 0$,

$$\operatorname{var}(F_{jk}) = \sigma_F{}^2 \; , \quad \operatorname{cov}(F_{jk}, F_{j'k'}) = 0$$

if

$$j' \neq j \quad \text{or} \quad k' \neq k \; , \quad \operatorname{cov}(C_{imn}, F_{jk}) = 0 \; ,$$

and

$$\begin{aligned} \operatorname{cov}(C_{imn}, C_{i'm'n'}) &= \alpha^{|n'-n|}\sigma_C{}^2 \; , && \text{if} \quad i'=i \quad \text{and} \quad m'=m \; , \\ &= 0 \; , && \text{if} \quad i' \neq i \quad \text{or} \quad m' \neq m \; . \end{aligned}$$

Here, $\sigma_F{}^2$, $\sigma_C{}^2$, and α are unknown parameters.

The correlation matrix of $C_{im1}, C_{im2}, \ldots$ is that for a first-order autoregressive process. The quantity C_{imn} can be interpreted as the deviation in the performance level of Team i in Week n of Year m from the level that is characteristic of Team i in Year m. The assumption (4.1) allows the weekly performance levels of any given team to be correlated to an extent that diminishes with elapsed time.

If the assumption (4.1) were adopted and positive values were used for α and $\sigma_C{}^2$, the effect on the statistical prediction algorithm would be an increased emphasis on the most recent of those games played in the year for which the prediction was being made. The games played early in that year would receive less emphasis.

The parameters $\sigma_T{}^2$, ρ, $\sigma_C{}^2$, α, and $\sigma_F{}^2$ associated with the modified model were actually estimated from the 1968–1976 NFL scores by an approximate restricted maximum likelihood procedure similar to that described in Section 2.3 for estimating parameters of the original model. As in Harville's (1977b) study of college football scores, there was no evidence that α differed from zero.

A second way to improve the accuracy of the statistical algorithm would be to supplement the information in the past scores with other quantitative information. A mixed linear model could be written for each additional variate. The random effects or the residual effects associated with each variate could be taken to be correlated with those for the other variates. At least in principle, the new variates could be incorporated into the prediction algorithm by following the same approach used in Section 2 in devising the original algorithm. In practice, depending on the number of variates that are added and the complexity of the assumed linear models, the computations could be prohibitive.

One type of additional variate would be the (signed) difference for each regular-season and playoff game between the values of any given statistic for the home team and the visiting team. For example, the yards gained by the home team minus the yards gained by the visiting team could be used. The linear model for a variate of this type could be taken to be of the same form as that applied to the difference in score.

There are two ways in which the incorporation of an additional variate could serve to improve the accuracy of the statistical prediction for the outcome S_{JK} of a future game. It could contribute additional information about the yearly characteristic performance levels $T_{h(J,K),J}$ and $T_{v(J,K),J}$ of the participating teams, or it could contribute information about the residual effect R_{JK}. Comparison of the estimates of σ_R^2 given in Table 1 with the figures given in Table 2 for average squared prediction error indicate that the first type of contribution is unlikely to be important, except possibly early in the season. Variates that quantify injuries are examples of variates that might contribute in the second way.

4.2 Other Approaches

The statistical prediction algorithm presented in Section 2 is based on procedures for mixed linear models described, for example, by Harville (1976, 1977a). These procedures were derived in a frequentist framework; however, essentially the same algorithm could be arrived at by an empirical Bayes approach like that described by Haff (1976) and Efron and Morris (1975) and used by the latter authors to predict batting averages of baseball players.

Essentially statistical algorithms for predicting the outcomes of NFL games were developed previously by Goode (as described in a nontechnical way by Marsha 1974), Rosner (1976), and Stefani (1977). Mosteller (1973) listed some general features that seem desirable in such an algorithm.

Comparisons of the results given in Section 3 with Stefani's results indicate that the predictions produced by the algorithm outlined in Section 2 tend to be more accurate than those produced by Stefani's algorithm. Moreover, Stefani reported that the predictions given by his algorithm compare favorably with those given by Goode's algorithm and with various other statistical and nonstatistical predictions.

There is some question whether it is possible for a bettor who takes an intuitive, essentially nonstatistical approach to beat the betting line (in the long run) more than 50 percent of the time. DelNagro (1975) reported a football prognosticator's claim that in 1974 he had made predictions for 205 college and NFL games relative to the betting line with 184 (89.8 percent) successes (refer also to Revo 1976); however, his claim must be regarded with some skepticism in light of subsequent well-documented failures (DelNagro 1977).

Winkler (1971) found that the collective rate of success of sportswriters' predictions for 153 college and NFL games was only 0.476, while Pankoff (1968), in a similar study, reported somewhat higher success rates.

Merchant (1973) followed the betting activities of two professional gamblers during the 1972 NFL season. He reported that they bet on 109 and 79 games and had rates of success of .605 and .567, respectively.

5. THE RATING PROBLEM

A problem akin to the football prediction problem is that of rating, ranking, or ordering the teams or a subset of the teams from first possibly to last. The rating may be carried out simply as a matter of interest, or it may be used to honor or reward the top team or teams.

The NFL has used what might be considered a rating system in picking its playoff participants. The NFL consists of two conferences, and each conference is divided into three divisions. Ten teams (eight before 1978) enter the playoffs: the team in each division with the highest winning percentage and the two teams in each conference that, except for the division winners, have the highest winning percentages. A tie for a playoff berth is broken in accordance with a complex formula. The formula is based on various statistics including winning percentages and differences in score for games between teams involved in the tie.

The prediction procedure described in Section 2 can also be viewed as a rating system. The ratings for a given year, say Year P, are obtained by ordering the teams in accordance with the estimates $\hat{T}_{1P}$, $\hat{T}_{2P}$, ... of their Year P characteristic performance levels. However, as a rating system, this procedure lacks certain desirable characteristics.

To insure that a rating system will be fair and will not affect the way in which the games are played, it should depend only on knowing the scores from the given season (Year P), should reward a team for winning per se, and should not reward a team for "running up the score" (Harville 1977b). The procedure in Section 2 can be converted into a satisfactory system by introducing certain modifications.

Define a "truncated" difference in score for Game k of Year P by

$$\begin{aligned} \dot{S}_{Pk}(M) &= M \ , & &\text{if} \quad S_{Pk} > M - \tfrac{1}{2} \ , \\ &= S_{Pk} \ , & &\text{if} \quad -(M - \tfrac{1}{2}) \le S_{Pk} \le M - \tfrac{1}{2} \ , \\ &= -M \ , & &\text{if} \quad S_{Pk} < -(M - \tfrac{1}{2}) \ , \end{aligned}$$

where M is some number of points. Let

$$\begin{aligned} &\check{S}_{Pk}(M; T_{h(P,k),P}, T_{v(P,k),P}) \\ &\quad = \tilde{S}_{Pk}(M; T_{h(P,k),P}, T_{v(P,k),P}; \hat{\sigma}_R^2, \hat{\lambda}, \hat{H}) \ , \end{aligned} \tag{5.1}$$

where $\tilde{S}_{Pk}(M; T_{h(P,k),P}, T_{v(P,k),P}; \sigma_R^2, \lambda, H)$ is the conditional expectation (based on an assumption that S_{P1}, S_{P2}, ..., T_{1P}, T_{2P}, ... are jointly normal) of S_{Pk} given $\dot{S}_{P1}(M)$, $\dot{S}_{P2}(M)$, ..., T_{1P}, T_{2P}, ..., that is, given all of the available truncated differences in score and all of the characteristic performance levels for Year P, or, equivalently, given $\dot{S}_{Pk}(M)$, $T_{h(P,k),P}$, and $T_{v(P,k),P}$. Note that

$$\check{S}_{Pk}(M; T_{h(P,k),P}, T_{v(P,k),P}) = S_{Pk} \quad \text{if} \quad |S_{Pk}| < M \ .$$

Let $\tilde{T}_{iP}(\lambda, H; S_{P1}, S_{P2}, \ldots)$ represent the conditional expectation of T_{iP} given S_{P1}, S_{P2}, ... In the modified rating procedure for Year P, the ratings are obtained by ordering estimates $\hat{T}_{1P}$, $\hat{T}_{2P}$, ..., where these estimates are based only on differences in score from Year P and where, instead of putting $\hat{T}_{iP} = \tilde{T}_{iP}(\hat{\lambda}, \hat{H}; S_{P1}, S_{P2}, \ldots)$

(as we would with the procedure in Section 2), we put

$$\hat{T}_{iP} = \bar{T}_{iP}(\hat{\lambda}, \hat{H}; \ddot{S}_{P1}(\hat{T}_{h(P,1),P}, \hat{T}_{v(P,1),P}), \ddot{S}_{P2}(\hat{T}_{h(P,2),P}, \hat{T}_{v(P,2),P}), \ldots) \quad (i = 1, 2, \ldots) \ . \tag{5.2}$$

Here, $\ddot{S}_{Pk}(\hat{T}_{h(P,k),P}, \hat{T}_{v(P,k),P})$ is an "estimated difference in score" given by

$$\beta\ddot{S}(1; \hat{T}_{h(P,k),P}, \hat{T}_{v(P,k),P}) + (1 - \beta)\ddot{S}(U; \hat{T}_{h(P,k),P}, \hat{T}_{v(P,k),P}) \ , \tag{5.3}$$

where U is some number of points (say $U = 21$) and β $(0 < \beta < 1)$ is some weight (say $\beta = \frac{1}{3}$) specified by the user.

Equations (5.2) can be solved iteratively for $\hat{T}_{1P}$, $\hat{T}_{2P}$, ... by the method of successive approximations. On each iteration, we first compute new estimates of T_{1P}, T_{2P}, ... by using the procedure in Section 2 with the current estimated differences in score in place of the "raw differences." Expressions (5.1) and (5.3) are then used to update the estimated differences in score.

The underlying rationale for the proposed rating system is essentially the same as that given by Harville (1977b) for a similar, but seemingly less satisfactory, scheme.

If the proposed rating system is to be accepted, it should be understandable to the public, at least in general terms. Perhaps (2.15) could be used as the basis for a fairly simple description of the proposed rating system. Our basic procedure is very similar to a statistical procedure developed by Henderson (1973) for use in dairy cattle sire selection. It is encouraging to note that it has been possible to develop an intuitive understanding of this procedure among dairy cattle farmers and to sell them on its merits.

It can be shown that taking $\hat{T}_{1P}$, $\hat{T}_{2P}$, ... to be as defined by (5.2) is equivalent to choosing them to maximize the function

$$\beta L_1(T_{1P}, T_{2P}, \ldots; \dot{S}_{P1}(1), \dot{S}_{P2}(1), \ldots; \hat{\sigma}_R^2, \hat{\lambda}, \hat{H}) + (1 - \beta)L_U(T_{1P}, T_{2P}, \ldots; \dot{S}_{P1}(U), \dot{S}_{P2}(U), \ldots; \hat{\sigma}_R^2, \hat{\lambda}, \hat{H})$$

where $L_M(T_{1P}, T_{2P}, \ldots; \dot{S}_{P1}(M), \dot{S}_{P2}(M), \ldots; \sigma_R^2, \lambda, H)$ is the logarithm of the joint probability "density" function of T_{1P}, T_{2P}, ..., $\dot{S}_{P1}(M)$, $\dot{S}_{P2}(M)$, ... that results from taking S_{P1}, S_{P2}, ..., T_{1P}, T_{2P}, ... to be jointly normal. When viewed in this way, the proposed rating system is seen to be similar in spirit to a system devised by Thompson (1975).

[*Received July 1978. Revised June 1979.*]

REFERENCES

DelNagro, M. (1975), "Tough in the Office Pool," *Sports Illustrated*, 43, 74–76.

——— (1977), "Cashing in a Sure Thing," *Sports Illustrated*, 47, 70–72.

Efron, Bradley, and Morris, Carl (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association*, 70, 311–319.

Haff, L.R. (1976), "Minimax Estimators of the Multinormal Mean: Autoregressive Priors," *Journal of Multivariate Analysis*, 6, 265–280.

Hartigan, J.A. (1969), "Linear Bayesian Methods," *Journal of the Royal Statistical Society*, Ser. B, 31, 446–454.

Harville, David A. (1976), "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," *Annals of Statistics*, 4, 384–395.

——— (1977a), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72, 320–338.

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

——— (1978), "Football Ratings and Predictions Via Linear Models," with discussion by Carl R. Morris, *Proceedings of the American Statistical Association*, Social Statistics Section, 74–82; 87–88.

——— (1979), "Recursive Estimation Using Mixed Linear Models With Autoregressive Random Effects," in *Proceedings of the Variance Components and Animal Breeding Conference in Honor of Dr. C.R. Henderson*, Ithaca, N.Y.: Cornell University, Biometrics Unit.

Henderson, Charles R. (1973), "Sire Evaluation and Genetic Trends," in *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, Champaign, Ill.: American Society of Animal Science, 10–41.

Marsha, J. (1974), "Doing It by the Numbers," *Sports Illustrated*, 40, 42–49.

Merchant, Larry (1973), *The National Football Lottery*, New York: Holt, Rinehart & Winston.

Mosteller, Frederick (1970), "Collegiate Football Scores, U.S.A.," *Journal of the American Statistical Association*, 65, 35–48.

——— (1973), "A Resistant Adjusted Analysis of the 1971 and 1972 Regular Professional Football Schedule," Memorandum EX-5, Harvard University, Dept. of Statistics.

Pankoff, Lyn D. (1968), "Market Efficiency and Football Betting," *Journal of Business*, 41, 203–214.

Patterson, H.D., and Thompson, Robin (1971), "Recovery of Inter-Block Information When Block Sizes Are Unequal," *Biometrika*, 58, 545–554.

Revo, Larry T. (1976), "Predicting the Outcome of Football Games or Can You Make a Living Working One Day a Week," in *Proceedings of the American Statistical Association*, Social Statistics Section, Part II, 709–710.

Rosner, Bernard (1976), "An Analysis of Professional Football Scores," in *Management Science in Sports*, eds. R.E. Machol, S.P. Ladany, and D.G. Morrison, Amsterdam: North-Holland Publishing Co., 67–78.

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

Thompson, Mark (1975), "On Any Given Sunday: Fair Competitor Orderings With Maximum Likelihood Methods," *Journal of the American Statistical Association*, 70, 536–541.

Winkler, Robert L. (1971), "Probabilistic Prediction: Some Experimental Results," *Journal of the American Statistical Association*, 66, 675–685.