An analysis of an ordinal-valued time series

David R. Brillinger

Department of Statistics, University of California, Berkeley, CA 94720

1 Abstract.

Time series and spatial processes are sometimes ordinal-valued. It can be convenient to handle such types of data via generalized linear model algorithms employing the complimentary *loglog* link function. This approach facilitates the use of standard statistical packages and leads to a convenient technique for handling serial dependence. Model fit is assessed by uniform residuals, amongst other tools. In this article an example of such an analysis is provided for a three-valued series corresponding to the possible results *loss*, *tie*, *win* of events involving a sports team.

> *"... - it is important to have in command the mathematics so you can solve the problem. Of course, the 64 dollar question is which mathematics to learn, because you can't learn all of it."*
>
> E.J. Hannan interviewed in [23]

2 Preamble.

Throughout my whole professional career Ted Hannan was there as a role model. The ever growing stack of his collected works was a constant research companion. In particular he was special for working on problems simultaneously from all sides: substantive, theoretical and computational. He always kept up with, indeed typically led, contemporary developments in time series. He has left us too soon, but his standards remain.

3 Introduction.

Ordinal data refers to quantities whose values are categories falling on a scale such that the order of the categories matters and is known. A characteristic is that adjacent categories may be sensibly merged with the ordinality remaining. One general reference is [21], Chapter 5. In the time

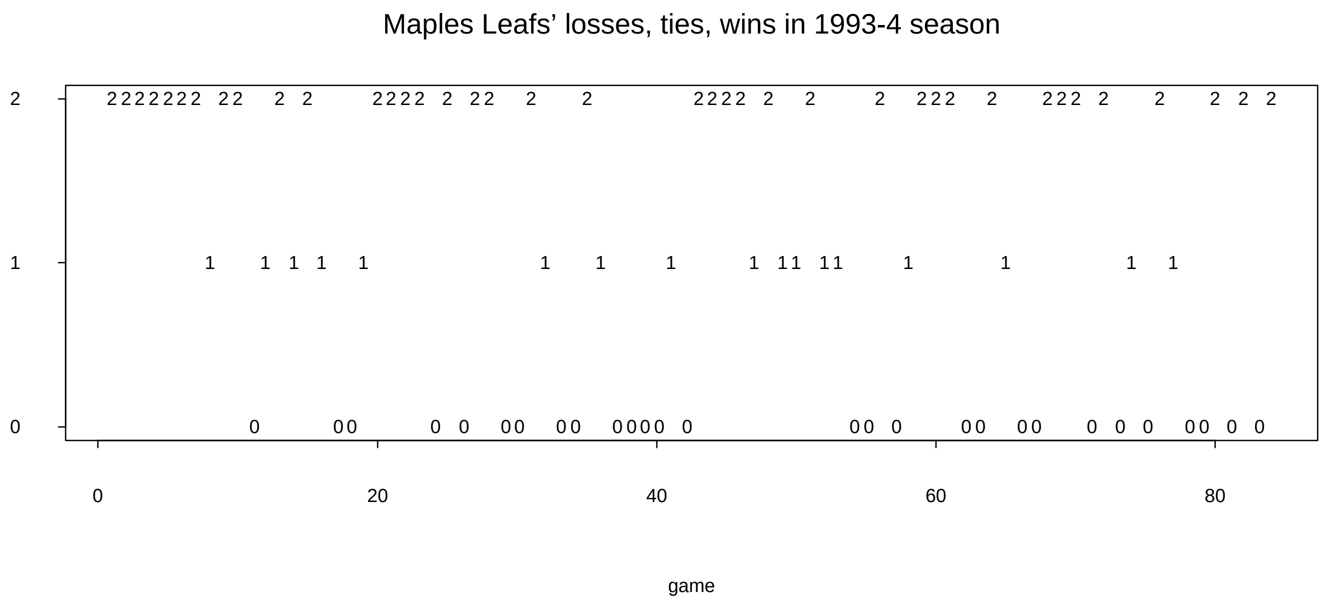


FIGURE 1. Results of 84 games: 0, 1, 2 refer to loss, tie, win respectively.

series case the individual values are ordinal categories and questions of interest include: Is there serial dependence? Is a trend present? Are there useful explanatories? Is there change?

This work is concerned with a segment $Y(t),\ t\ =\ 0, ..., T-1$ of an ordinal-valued time series taking values such that

$$Y(t)\ =\ 0,\ 1,\ or\ 2$$

corresponding to the results *loss*, *tie*, *win* of the Toronto Maple Leafs Hockey team during the 1993-1994 season. (In assigning points in the standings 0, 1, 2 actually represent the points awarded.)

Figure 1 provides a graph of the results. There were 84 regular season games in all. The Toronto team began the season with a record setting string of 10 wins. In order that the results be more homogeneous for the analyses presented, the data graphed and analysed actually correspond to the state of the game after regulation time. (If the game is tied at the end of regulation time, an overtime is period is played, which may result in a win for one of the teams.) By this count Toronto had 28 losses, 17 ties and 39 wins in the course of the season.

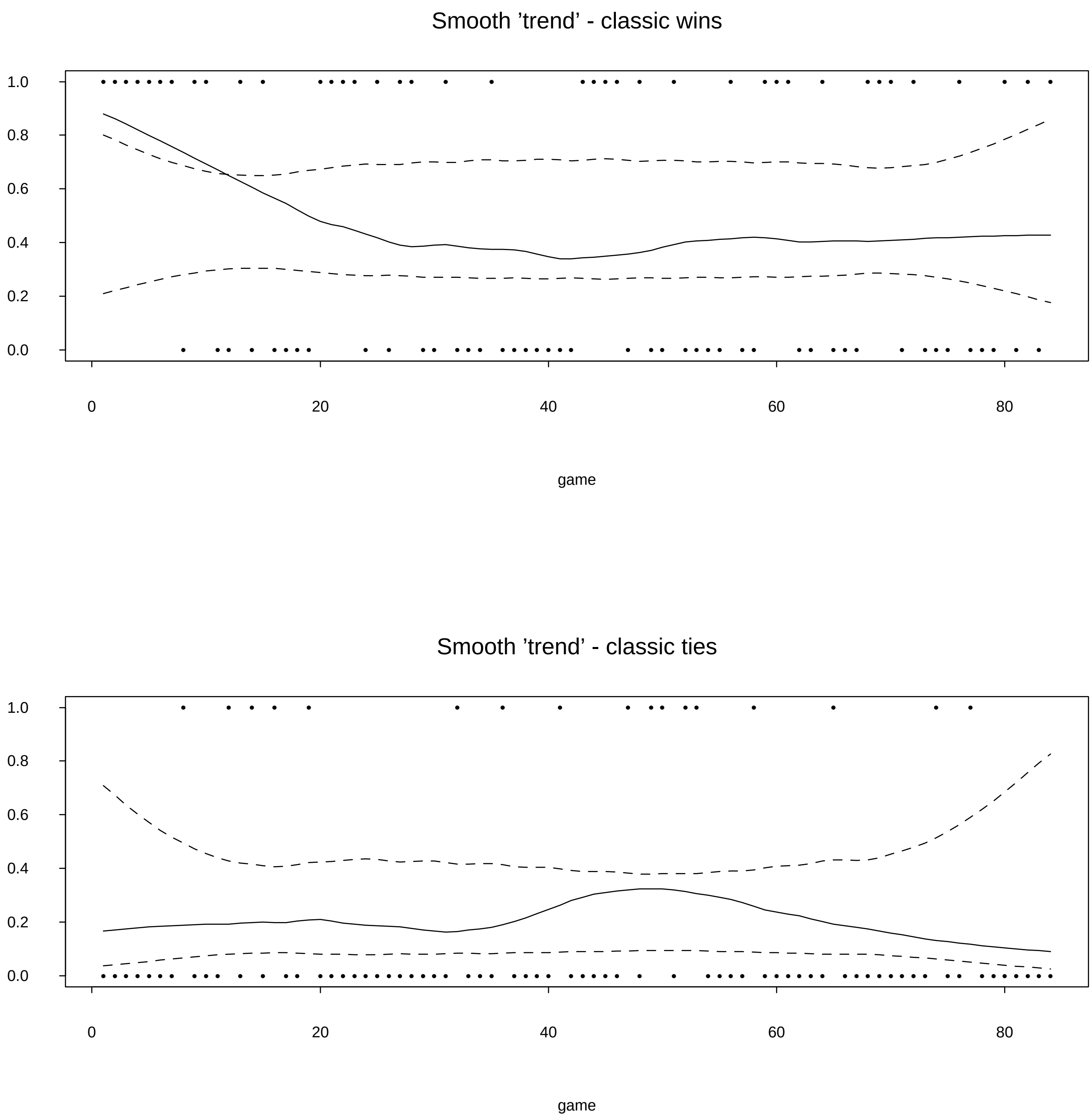


FIGURE 2. Smoothed rate for wins and ties respectively with uncertainty limits.

Figure 2 provides smoothed estimates of the probability of a classic (i.e. after regulation time) win and of a classic tie respectively. The approximate ± 2 standard error limits are computed as if the successive games are statistically independent. Except for the early success, the win curve fluctuates about a constant mean level. In the case of the ties the curve fluctuates about the mean level throughout. These curves were produced employing the *cloglog* link and the functions *gam*() and *predict.gam*() of the statistical package S (see[2, 8]).

The data are provided in an Appendix.

4 Ordinal Data.

A number of different models have been proposed for the analysis of ordinal data. These include: continuation ratio (see [12]), stereotype (see [1]) and the grouped continuous (see [20]).

The following presents an approach to building a stochastic model for ordinal data. Let Y be 0, 1, 2 for a particular game, depending on whether the result is a *loss*, *tie* or *win*. Suppose that there exists a latent or state variable, Λ, whose value represents the strength of the Toronto team against a general opponent. Assume the existence of cutpoints a and b such that

$$Y = 0 \; if \; \Lambda < a, \quad Y = 1 \; if \; a < \Lambda < b \; and \quad Y = 2 \; if \; b < \Lambda$$

So for example

$$Prob\{Y = 1\} = F_\Lambda(b) - F_\Lambda(a) \tag{0.1}$$

where F_Λ is the c.d.f. of Λ. Figure 3 presents an example of a graph of a possible density function for Λ with the regions of *loss*, *tie*, *win* indicated. In the graph the term linear predictor refers to Λ. The approach involving a latent variable has the advantages of: easy interpretability, clear possibilities of merging adjacent categories and of flexibility.

Maximum likelihood is a natural method of estimating unknown parameters in many cases and will be employed in the present work. Goodness of fit may be assessed by procedures such as: deviance and chi-squared type statistics (see [21]), plots of estimated probability against the linear predictor ([5, 6]) or "uniform residuals" ([4]).

In a generalized linear model, the link function describes the relation between the mean of the basic variate and the natural parameter of its distribution. Its choice is sensibly based on the subject matter of the problem. The complimentary *loglog* corresponds to situations in which of an extremal variate crosses a threshold and an extreme value distribution, ([25]). In the present context this may be reasonable, with a win for the hockey team resulting from the team members putting out maximum efforts to exceed those of the opponent.

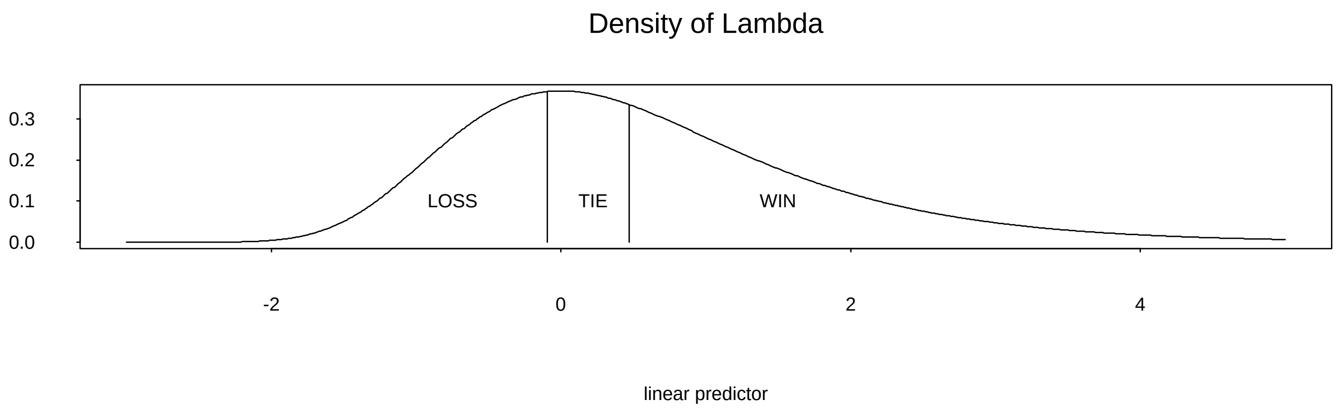


FIGURE 3. Areas of regions refer to probabilities of the respective events.

The extreme value distribution of the first type is given by

$$Prob\{\Lambda \ > \lambda\} \ = \ exp\{-e^{\lambda}\} \quad for \ \lambda \ > \ 0$$

The graph of Figure 3 is based on this distribution. One can write

$$log(-log(1 - Prob\{\Lambda \leq \lambda\})) \ = \ \lambda$$

and sees the appearance of the *cloglog* link. In the case of ordinal-valued Y one writes

$$log(-log(1 - Prob\{Y \leq j\} \ = \ \theta_j$$

with $\theta_j \ > \ \theta_{j-1}$ and

$$log(-log(1 - Prob\{Y \ = \ j \mid Y \ \geq \ j\})) \ = \ \alpha_j \qquad (0.2)$$

for $j \ = \ 0, \ 1, \ 2$. Pregibon, [24], noted the fact that for the *cloglog* link the parametrization was of the same form and hence, by writing a probability as a product of conditional probabilities, one could employ standard statistical packages in analyses of such multinomial data. See also [17]. One can work with $Prob\{win\}$ and $Prob\{tie|not \ win\}$ in the present hockey game case.

Explanatory variables, x, may be introduced quite directly by writing

$$\Lambda \ = \ \mathrm{E} \ + \ \beta' x$$

where E has the extreme value distribution. Now (0.1) becomes

$$F_{\mathrm{E}}(b - \beta' x) \ - \ F_{\mathrm{E}}(a - \beta' x)$$

5 The Time Series Case.

There is a massive literature concerning time series, that is sequences $Y(t)$, $t = 0, \pm 1, \pm 2, \ldots$ which are stochastic. The literature mainly refers to real-valued Y, some of it refers to count-valued ([3, 14, 18, 28]). What distiguishes the present circumstance are the values that Y can take on. In this work the values correspond to ordinal categories. In the case of two categories the series are binary and there is a large existing literature ([5, 9]). There is further a literature for extensions to the case of the generalized linear models ([11, 10, 15, 16, 26, 27]). There are also approaches to categorical-valued time series based on Markov chains and on state space descriptions ([11, 10]). A distinction that arises in the literature concerns whether one realization of the time series is involved or several. The latter case is typically referred to as longitudinal data analysis ([7, 19, 22]).

Both parametric and nonparametric models can be considered. A direct parametric way to introduce temporal dependence is to set up an autoregressive-type model with past values of the series being employed as explanatories. In likelihood approaches it is then convenient to set up a likelihood as the product of a sequence of conditional mass or density functions, f_Y,

$$\prod_{t=0}^{T-1} f_{Y(t)}(y_t|H_{t-1})$$

with H_t denoting the history $\{y_0, \ldots, y_t\}$. Taking this result together with the simplification resulting from the use of the complimentary *loglog* function, referred to in Section 4, means that parametric analyses can be carried out using standard functions such as *glm*() of S, [8]. The appearance of the conditional term (0.2) may be controlled by the use of the weight option.

6 Results

The graphs of Figure 1 may be considered a first-order analysis of the question of temporal dependence. What may be seen is a small indication of an increased probability of a win for the Toronto team at the beginning of the season. It is of further interest whether there is some clustering of the *losses*, *ties* or *wins* or if these perhaps alternate in some fashion.

A nonparametric second-order analysis may be developed by creating a bivariate time series. Define the two binary series Y_1 and Y_2 with $Y_1(t) = 1$ if the t-th game is a *win* and 0 otherwise, similarly define $Y_2(t) = 1$ if the game is a *tie* and 0 otherwise. To begin consider a frequency domain approach, the one so often taken by Ted Hannan [13]. In the case of a bivariate stationary white noise process, each of the second-order spectra are constant and the quadrature spectrum is identically 0. Figure 4 provides

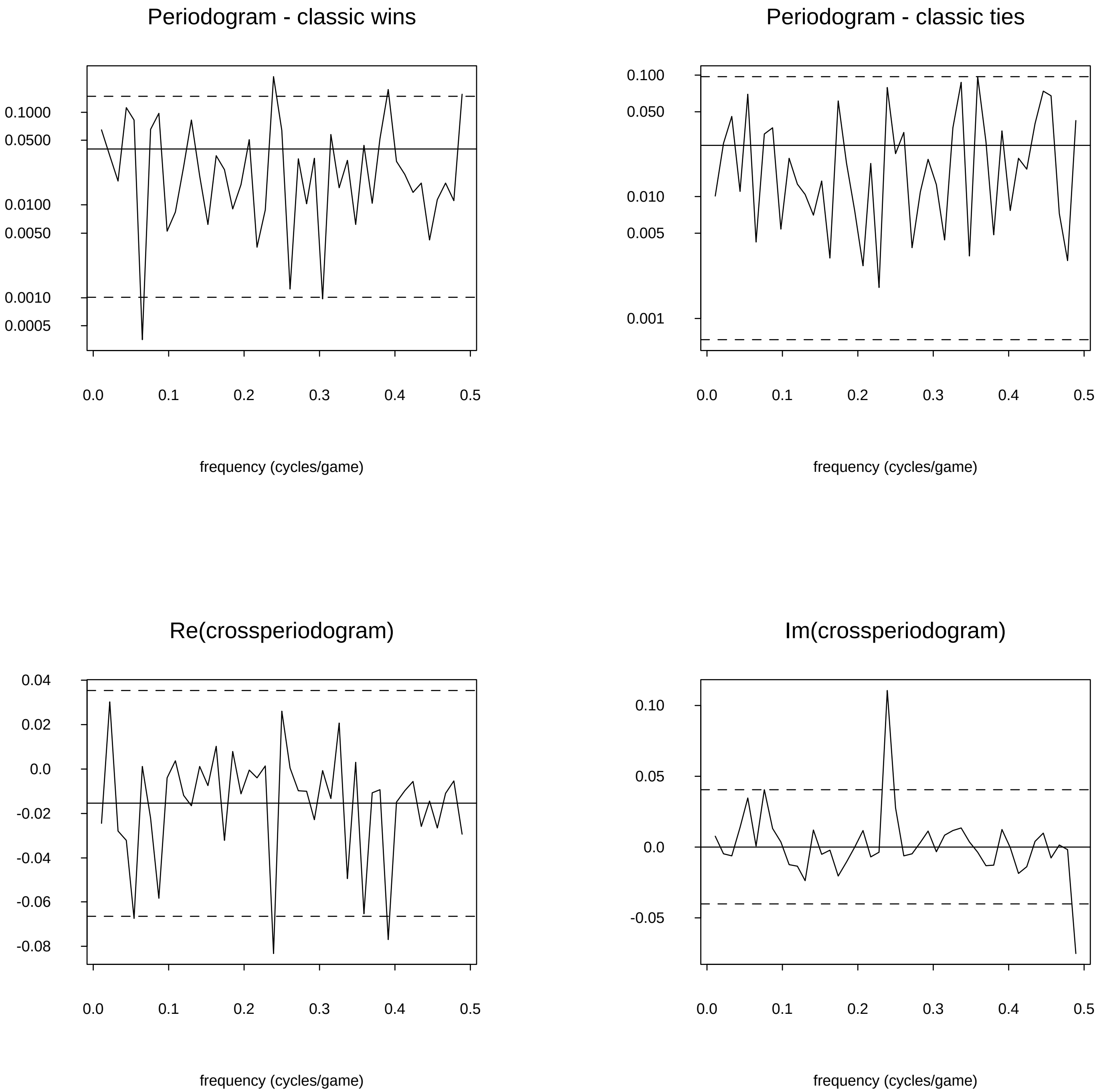


FIGURE 4. Second-order periodograms of the data.

the periodograms and cross-periodograms of the data for the series Y_1 and Y_2. The solid lines are the estimated levels in the case that the successive observations are i.i.d. The dashed lines are approximate 95% marginal confidence limits. There is one unusual point in the crossperiodogram, but no substantial evidence for temporal dependence.

One type of parametric analysis involves fitting a process of autoregressive type. As an example consider the model

$$log(-log(1 - Prob\{Y(t) = j|H_{t-1}\})) = \theta_j + \phi_j \cdot y_{t-1} \qquad (0.3)$$

with the $\phi \cdot y$ term having the meaning that the value, y_{t-1}, of the series at the previous time point is to be viewed as a factor. The deviance change in fitting the model 0.3 with and without this term is 3.03 on 4 degrees of freedom with a corresponding probvalue of .553. There is no evidence for the postulated form of dependency on the previous time value. Earlier time values may be studied just as easily.

Various other explanatories may be considered, for example whether the game is home or away, goals scored and some measure of the strength of the opposing team. In the case of including whether the game was home or away, as an explanatory factor, the deviance change is only .012 on 1 degree of freedom. The corresponding probvalue is .911 . Again there is no evidence of an effect.

7 Goodness of Fit.

In any work with stochastic models, goodness of fit is a central issue. In work with generalized linear models the residual deviance is often employed; however its approximation by a chisquared variate in the null case is often poor. In [4] the idea of employing uniform residuals was introduced. One uses the probability integral transformation based on the fittted model. In the case that the parameter values are known, this will have a uniform distribution. These residuals may be plotted against explanatories, be used to construct probability plots and other such things.

The present approach acts as if the data are binary. Suppose that X is a Bernoulli variate with $Prob\{X = 1\} = \pi$. Then a standard uniform variate, V, may be constructed by setting

$$V = uniform\ on\ (1-\pi, 1)\ if\ X = 1$$

$$V = uniform\ on\ (0, 1-\pi)\ if\ X = 0$$

This was done for the simplest model (of the Y i.i.d.) and the observed data, based on the estimates of $Prob\{win\}$ and $Prob\{tie|not\ win\}$. Figure 5 gives plots of the V's against game for the wins and conditional ties. In

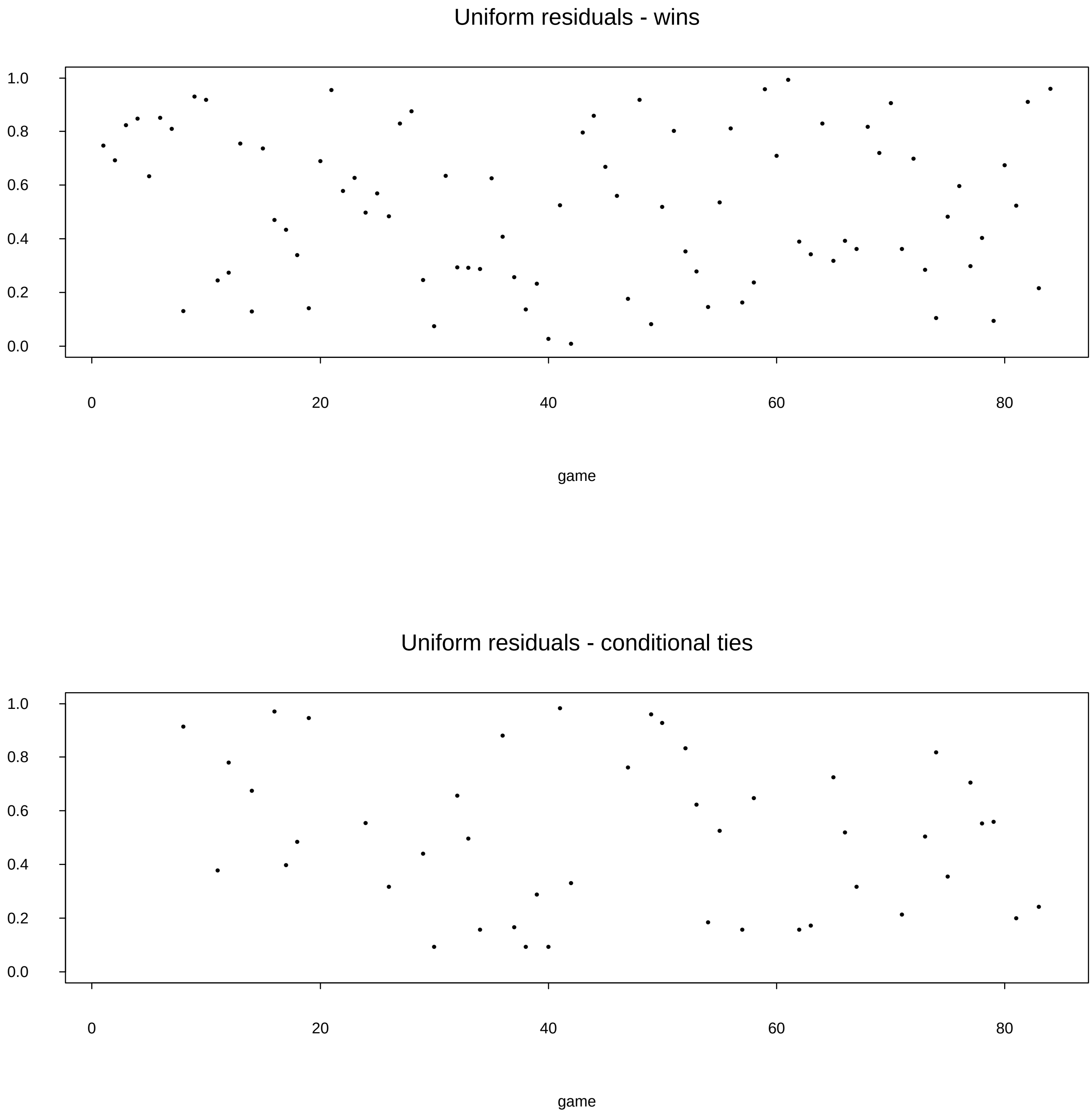
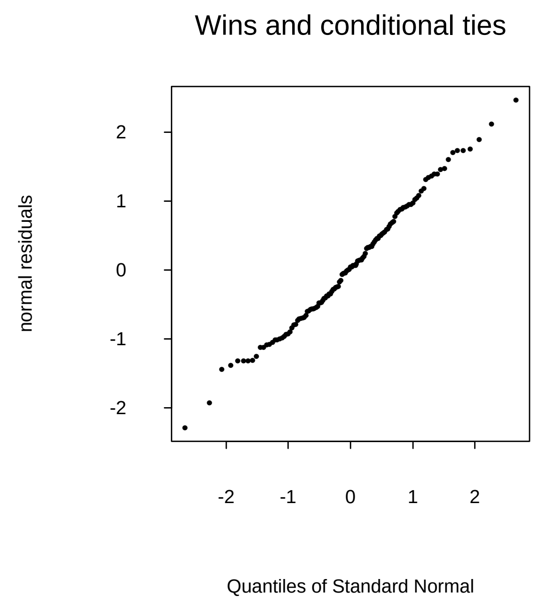


FIGURE 5. Variates created to be approximately standard uniform if the model holds.



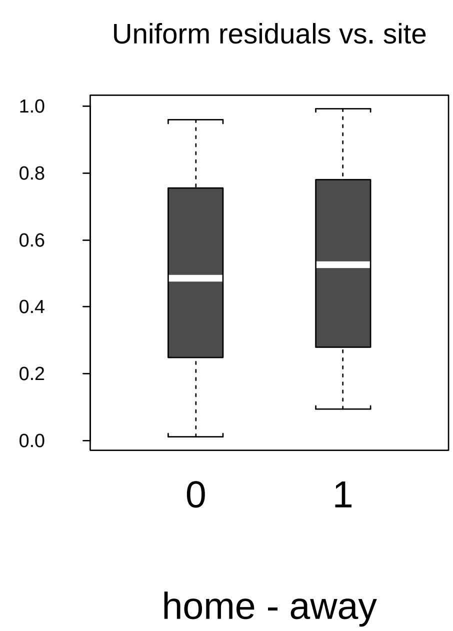


FIGURE 6. In top figure the uniform variates have been transformed to normals. In the lower boxplots of uniform residuals are plotted in the home and away cases.

the first case one sees some elevated values at the beginning, but randomness therafter. In the second case there is apparent randomness. Figure 6 gives a normal probability plot and a plot against the home-away variate respectively. There is no evidence to contradict the assumptions of the fitted model.

8 Summary.

The 1993-94 Toronto team began the season with a string of successes; however ultimately the results of the various games appear random. The analyses provide no real evidence for temporal dependence in rate or serial correlation. If temporal dependence had been noted there would have been the possibility of using the model for prediction.

9 Acknowledgements

This work was carried out with the partial support of the National Science Foundation Grant DMS-9300002 and the Office of Naval Research Grant N00014-94-1-0042. Mark Rizzardi made some helpful remarks on the manuscript.

10 Appendix

The site variable refers to whether the game was home or away, 0 is away. The overtime variable refers to whether the game ended in regulation time, 0 means it did.

Goals for	Goals against	site	overtime
6	3	0	0
2	1	0	0
5	4	1	0
7	1	0	0
6	3	0	0
2	1	1	0
7	2	0	0
4	3	1	1
2	0	1	0
4	2	1	0
2	5	1	0
3	3	1	1
6	3	0	0
3	3	1	1
5	3	0	0
2	2	1	1
2	3	1	0
2	3	0	0
5	5	0	1
4	3	1	0
3	2	1	0
3	2	1	0
5	2	1	0
3	5	1	0
4	2	0	0
0	3	0	0
4	2	0	0
5	4	1	0
3	4	0	0
4	5	0	0
3	1	0	0
3	3	1	1
0	1	0	0
2	6	1	0
4	1	0	0
2	2	0	1
2	3	1	0
2	5	1	0
0	4	1	0
4	7	0	0
3	3	1	1
0	1	0	0

Goals for	Goals against	site	overtime
6	3	0	0
5	3	0	0
3	0	1	0
2	1	1	0
4	3	0	1
5	1	1	0
3	3	0	1
3	3	1	1
4	3	0	0
4	4	0	1
4	4	1	1
3	4	0	0
1	2	0	0
3	1	1	0
2	3	1	0
5	4	0	1
2	1	0	0
3	2	0	0
6	4	1	0
3	6	1	0
0	3	0	0
4	1	1	0
6	5	1	1
1	4	1	0
2	3	0	0
4	2	0	0
4	2	1	0
3	1	0	0
1	4	0	0
4	2	0	0
3	6	0	0
1	1	1	1
1	2	0	0
6	3	0	0
2	3	1	1
3	5	1	0
1	3	1	0
6	4	1	0
3	5	1	0
7	0	0	0
3	4	0	0
6	4	1	0

11 References

[1] J.A. Anderson. Regression and ordered categorical variates. *J. Royal Statist. Soc. B*, 46:19–35, 1984.

[2] R.A. Becker, J.M. Chambers, and A.R. Wilks. *The New S Language.* Wadsworth, Pacific Grove, 1988.

[3] D.R. Brillinger. The natural variability of vital rates and associated statistics. *Biometrics*, 42:693–734, 1986.

[4] D.R. Brillinger and H.K. Preisler. Maximum likelihood estimation in a latent variable problem. In S. Karlin et al., editor, *Studies in Econometrics, Time Series and Multivariate Statistics*, pages 31–65, New York, 1983. Academic.

[5] D.R. Brillinger and J.P. Segundo. Empirical examination of the threshold model of neuron firing. *Biol. Cybernetics*, 35:213–220, 1979.

[6] D.R. Brillinger, A. Udias, and B.A. Bolt. A probability model for regional focal mechanism solutions. *Bull. Seismol. Soc. Amer.*, 70:149–170, 1980.

[7] G.J. Carr, K.B. Hafner, and G.G. Koch. Analysis of rank measures of association for ordinal data from longitudinal studies. *J. Amer. Statist. Assoc.*, 84:797–804, 1989.

[8] J.M. Chambers and T.J. Hastie. *Statistical Models in S.* Wadsworth, Pacific Grove, 1992.

[9] D.R. Cox. *Analysis of Binary Data.* Methuen, London, 1970.

[10] L. Fahrmeir. State space modelling and conditional mode estimation for categorical time series. In D.R. Brillinger et al., editor, *New Directions in Time Series*, pages 87–110, New York, 1992. Springer.

[11] L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer-Verlag, New York, 1994.

[12] S.E. Fienberg. *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, Mass., 1980.

[13] E.J. Hannan. *Multiple Time Series.* Wiley, New York, 1970.

[14] A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge Press, Cambridge, 1989.

[15] H. Kaufmann. Regression models for nonstationary categorical time series: asymptotic estimation theory. *Ann. Statist.*, 15:79–98, 1987.

[16] G. Kitagawa and W. Gersch. *Smoothness Priors Analysis of Time Series*. Springer Lecture Notes, 1995.

[17] E. Laara and J.N.S. Matthews. The equivalence of two models for ordinal data. *Biometrika*, 72:206–207, 1985.

[18] A. Latour. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Preprint*, 1995.

[19] K.Y. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

[20] P. McCullagh. Regression model for ordinal data. *J. Roy. Statist. Soc. B*, 42:109–127, 1980.

[21] P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall, New York, 1989.

[22] R.D. Murison. Analysis of repeated measure of ordinal data. *Proc. Centre Math. Applications, Australian National University*, 28:109–118, 1991.

[23] A. Pagan. The ET interview: Professor E.J. Hannan. *Econometric Theory*, 1:263–289, 1985.

[24] D. Pregibon. Discussion of paper by P. McCullagh. *J. Royal Statist. Soc. B*, 42:139–139, 1980.

[25] H.K. Preisler. Analysis of a toxicological experiment using a generalized linear model with nested random effects. *Int. Statist. Review*, 57:145–159, 1989.

[26] M. West, P.J. Harrison, and H.S. Mignon. Dynamic generalized linear models and bayesian forecasting. *J. Amer. Statist. Assoc.*, 80:73–97, 1985.

[27] M. West, P.J. Harrison, and H.S. Migon. *Bayesian Forecasting and Dynamic Models*. Springer, New York, 1989.

[28] S.L. Zeger. A regression model for time series of counts. *Biometrika*, 75:621–629, 1988.