Nonparametric additive regression models for binary time series

Rob J. Hyndman, Monash University March 1999

Abstract: I consider models for binary time series, starting with autoregression models and then developing generalizations of them which allow nonparametric additive covariates. I show that several apparently different binary AR(1) models are equivalent. Three possible nonparametric additive regression models which allow for autocorrelation are considered; one is a generalization of an ARX model, the other two are generalizations of a regression model with AR errors. One of the models is applied to two data sets: IBM stock transactions and Melbourne's rainfall. The fitted models show that stock transaction occurrences are more likely if there have been large transactions in the previous time period. They also show that the Southern Oscillation Index does not provide a strong predictor of rainfall occurrence in Melbourne, contrary to current meteorological practice.

Key words: ARX models, autocorrelated errors; autocorrelation; binary time series; generalized additive model; generalized linear model; logistic regression; non-Gaussian time series, smoothing with correlated errors; time series regression.

JEL classification: C14, C22, C25.

Author contact details:

Address:	Department of Econometrics and Business Statistics, Monash University, Clayton VIC 3168, Australia.
Telephone:	(03) 9905 2358.
Fax:	(03) 9905 5474.
Email:	Rob.Hyndman@buseco.monash.edu.au.

1 Introduction

Binary time series arise whenever the occurrence of an event is of interest. For example the occurrence of sales for slow-moving manufactured goods subject to intermittent demand, or the occurrence of transactions on a heavily traded stock in a short time interval.

Models for cross-sectional binary data have been studied extensively and include probit and logistic regression models (see McCullagh and Nelder, 1989). Recently, these have been extended to additive logistic models which allow smooth, non-linear covariate relationships to be estimated nonparametrically (see Hastie and Tibshirani, 1990).

However, the modelling of binary time series data has not received much attention, and nonparametric additive regression models with autocorrelation has not previously been considered to my knowledge.

In Section 2 I define an autoregressive model for binary time series and compare it with other binary time series models which have been proposed. It is shown that the first order model for a number of these proposals is equivalent to the model defined here.

The extension of the AR model to incorporate smooth (non-linear) additive regression terms is considered in Section 3. Three possible additive regression models which allow for autocorrelation are considered. I show that a generalization of an ARX model provides easy marginal interpretation of the effect of covariates and inference can be obtained using the techniques developed for generalized additive models (GAMs).

The rest of the paper concerns two examples in which a nonparametric additive regression model is applied to real data. Section 4 considers models of the occurrence of IBM stock transactions in one-minute intervals and Section 5 considers models of the daily occurrence of rainfall in Melbourne using a number of covariates.

2 Binary autoregressive models

Define a binary AR(p) process to be the two-state Markov chain $\{Y_t\}$ on $\{0, 1\}$ with $t = 0, 1, 2, \ldots$, and transition probabilities

$$\Pr(Y_t = 1 \,|\, \boldsymbol{Y}_{t-1}) = \ell^{-1}\left(\lambda + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}\right) \tag{2.1}$$

where $\boldsymbol{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_0)'$ and ℓ denotes a link function. Two important cases are the identity link function $\ell(u) = u$ and the logistic link function given by

$$\ell(u) = \log\left(\frac{u}{1-u}\right). \tag{2.2}$$

Cox and Snell (1989) consider model (2.1) with the logistic link function. In this case, $\ell^{-1}(u) = e^u/(1+e^u)$ so that the parameters $\lambda, \phi_1, \ldots, \phi_p$ may take any real values. However, it is difficult to establish autocovariance properties of the process using a logistic link function.

For the remainder of this section, we will assume an identity link function. In this case, some parameter restrictions are necessary to ensure the transition probabilities defined by (2.1) lie between 0 and 1. Let $\{\phi_1^*, \ldots, \phi_K^*\}$ denote the sums of all subsets of $\{\phi_1, \ldots, \phi_p\}$ and let

$$\phi_{\max} = \max\{\phi_1^*, \ldots, \phi_K^*\} \qquad \text{and} \qquad \phi_{\min} = \max\{\phi_1^*, \ldots, \phi_K^*\}.$$

Then the parameters λ and $\phi_1, \ldots, \phi_p$ can take any values such that

$$0 < \lambda < 1 \qquad 0 < \lambda + \phi_{\max} < 1 \qquad \text{and} \qquad 0 < \lambda + \phi_{\max} < 1. \tag{2.3}$$

Under these conditions, the Markov chain is clearly irreducible, aperiodic and positive recurrent. Therefore it is ergodic with steady state probabilities

$$\Pr(Y_t = 1) = \mu = \lambda/(1 - \phi_1 - \cdots - \phi_p) \qquad \text{and} \qquad \Pr(Y_t = 0) = 1 - \mu,$$

and the stationary distribution of Y_t is Bin(μ).[1]

The conditional mean of the process is linear

$$\mathrm{E}(Y_t \mid \boldsymbol{Y}_{t-1}) = \lambda + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} \tag{2.4}$$

and so the model satisfies the definition of an autoregressive process given by Grunwald et al. (1997).

One consequence of the linear conditional mean (2.4) is the set of Yule-Walker equations

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \cdots + \phi_p\gamma(h-p), \qquad h = 1, 2, \ldots, \tag{2.5}$$

where $\gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h})$. For $h = 0$ we have (from the stationary distribution)

$$\gamma(0) = \mu(1 - \mu).$$

In the case $p = 1$, these lead to the familiar geometrically decaying autocorrelation function:

$$\rho(h) = \mathrm{Corr}[Y_t, Y_{t-h}] = \phi_1^h, \qquad h = 1, 2, \ldots. \tag{2.6}$$

2.1 Alternative formulations

The binary AR(1) model (2.1) is fully specified by two parameters. Consequently, any irreducible aperiodic Markov process defined on $\{0, 1\}$ with two free parameters will be equivalent except for some possible additional parameter restrictions.

An early proposal for binary autoregression was that of Jacobs and Lewis (1978a,b) who defined the first order process

$$Y_t = \begin{cases} Y_{t-1} & \text{w.p. } \phi \\ Z_t & \text{w.p. } 1 - \phi, \end{cases} \tag{2.7}$$

[1] The notation Bin(π) denotes a Bernoulli distribution with probability π of taking value 1 and probability $1 - \pi$ of taking value 0.

where $Z_t \sim \text{Bin}(\lambda/(1-\phi))$ is an iid sequence of random variables. There are two free parameters and so this formulation is equivalent to the binary AR(1) process although it restricts ϕ to be positive. Jacobs and Lewis (1978a,b) and Lewis (1980) generalize this model to allow an ARMA(p,q) structure. However, they set $\phi_1 = \phi_2 = \phi_{p-1} = 0$ for $p > 1$ and so they do not obtain the binary AR(p) defined by (2.1).

McKenzie (1985) proposed the first order process

$$Y_t = \phi_t Y_{t-1} + Z_t \tag{2.8}$$

where $\phi_t = A_t - Z_t$, $A_t \sim \text{Bin}(\lambda+\phi)$ and $Z_t \sim \text{Bin}(\lambda)$. Again, there are two free parameters and so this definition is equivalent to the binary AR(1) process. It is also equivalent to the model of Jacobs and Lewis for $0 < \phi < 1$ (an equivalence which appears to have been overlooked by McKenzie). McKenzie mentions a generalization of the formulation (2.8) to higher order autoregression but gives no details.

Kanter (1975) proposes a first order model which has the autocorrelation function (2.6) but not the linear conditional mean (2.1). This model is much more difficult to generalize to allow regressors and I do not consider it further.

Another approach is to assume that the discrete process inherits some structure from a related underlying continuous process. This approach is explored by Lomnicki and Zaremba (1955), Kedem (1980a,b) and Keenan (1982).

3 Binary regression with autocorrelation

There are several ways of generalizing the binary AR(p) model to allow explanatory variables. In this section I shall outline three possible model formulations. These are analogues of the two main methods for modelling Gaussian data with regressors and autoregression, namely ARX models and regression models with autoregressive errors.

Let $\boldsymbol{X}_t = (X_t, \ldots, X_{t-p})'$, and $\boldsymbol{D}_t' = \{\boldsymbol{X}_t', \boldsymbol{Y}_{t-1}'\}$, for $t = p+1, \ldots, n$.

An additive ARX model for Gaussian data can be written as $Y_t \mid \boldsymbol{D}_t \sim \text{N}(m_t, \sigma^2)$ where

$$m_t = \lambda + \sum_{i=1}^{r} g_i(X_{i,t}) + \sum_{j=1}^{p} \phi_j Y_{t-j}, \tag{3.1}$$

where each g_i is a smooth (not necessarily linear) function of the covariates $X_{i,t}$.

A Gaussian additive model with autocorrelated errors is given by

$$Y_t = \lambda + \sum_{i=1}^{r} g_i(X_{i,t}) + \varepsilon_t \tag{3.2}$$

where the errors follow a Gaussian AR(p) process

$$\varepsilon_t = \sum_{j=1}^{p} \phi_j \varepsilon_{t-j} + \eta_t\,, \qquad \eta_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)\,. \tag{3.3}$$

This model is considered by Niu (1996). If $r = 1$, it is the model for smoothing with autoregressive errors used by Diggle and Hutchinson (1989), Altman (1990) and Hart (1991).

Rewrite (3.2) as a conditional distribution $Y_t \mid \boldsymbol{D}_t \sim \mathrm{N}(m_t, \sigma^2)$ with

$$m_t = \mu_t + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu_{t-j}), \tag{3.4}$$

where

$$\mu_t = \lambda + \sum_{i=1}^{r} g_i(X_{i,t}). \tag{3.5}$$

Note that $\mathrm{E}(Y_t \mid \boldsymbol{X}_t) = \mu_t$ for the model (3.2) and (3.3). So each g_i can be can be interpreted in terms of the effect on the marginal mean μ_t. With ARX models, each g_i represents the effect of the variable $X_{i,t}$ on the mean of Y_t conditional on the values of the other covariates and on the values of $Y_{t-1}, \ldots, Y_{t-p}$ and even on the order p. It is difficult to formulate any marginal interpretation of the effect of the covariates for this model. Consequently, the model (3.2) and (3.3) is often preferred over the ARX model (3.1).

In the following sections we shall see that the opposite is true for binary models. With binary data, the analogue of the ARX model provides a simple marginal interpretation whereas the analogue of the regression with AR errors has either no obvious marginal interpretation or difficulties with the domain of the conditional mean.

3.1 Transitional binary additive models

A natural analogue of the Gaussian ARX model (3.1) is to let $Y_t \mid \boldsymbol{D}_t \sim \mathrm{Bin}(m_t)$ where

$$\ell(m_t) = \lambda + \sum_{i=1}^{r} g_i(X_{i,t}) + \sum_{j=1}^{p} \phi_j Y_{t-j} \tag{3.6}$$

and ℓ is a link function. I call this model a **transitional binary additive model**.

When $r = 0$ and ℓ denotes the identity function, we obtain the binary AR(p) model specified by (2.1).

When each g_i is linear, the model is a transitional generalized linear model (GLM) and has been discussed by Diggle, Liang and Zeger (1994, chapter 10). More generally, the model is a binary additive model with lagged values of the response variable as covariates. Thus, it is within the class of generalized additive models (Hastie and Tibshirani, 1990). If each g_i is linear, it falls with the class of generalized linear models (McCullagh and Nelder, 1989).

Estimation of (3.1) is possible using Hastie and Tibshirani's backfitting algorithm in conjunction with a nonparametric smoothing method such as locally weighted polynomials (Fan and Gijbels, 1996), smoothing splines (Green and Silverman, 1994) or penalized regression splines (Eilers and Marx, 1996) for each of the smooth terms g_i. This is easily accomplished using standard GAM software (such as the `gam` function in S-Plus)

Because the terms in Y_{t-k} $(k = 1, \ldots, p)$ contribute only constants to the transformed mean of Y_t, g_i can be interpreted as the effect (up to a constant) of variable $X_{i,t}$ on the link-transformed mean of Y_t conditional only on the values of the other covariates. This marginal interpretation of g_i is not possible with the Gaussian model (or with other non-Gaussian models); it is a fortunate consequence of the sample space $\{0, 1\}$ for Y_t.

However, one problem with this model is that the autocorrelation structure of the errors is lost because of the need to use a logistic (or similar) link function to restrict the transition probabilities to lie between 0 and 1. For example, in the simplest case, $p = 1$, $\text{Corr}(Y_t, Y_{t-h} \mid \boldsymbol{X}_t) \neq \phi^h$ unless the identity link is used.

3.2 Transitional binary additive models with lagged covariates

The obvious binary analogue of (3.2) is

$$Y_t \mid \boldsymbol{D}_t \sim \text{Bin}(m_t) \qquad \text{where} \qquad \ell(m_t) = \mu_t + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu_{t-j}) \tag{3.7}$$

where μ_t is given by (3.5) and ℓ is a link function.

When $r = 0$ and ℓ denotes the identity function we obtain the binary AR(p) model specified by (2.1).

If each of the g_i functions is linear, (3.7) is equivalent to the binary Markov regression models proposed by Zeger and Qaqish (1988) and also considered by Li (1991, 1994). Grunwald and Hyndman (1998) consider (3.7) with $g = \ell$ the identity function, $r = 1$ and $X_{1,t} = t$.

Although this model is the most natural formulation in the Gaussian setting, it often has difficulties of interpretation with non-Gaussian series. Consider the case $p = 1$ so that $\text{E}(Y_t \mid D_t) = \ell^{-1}(\mu_t + \phi_1 Y_{t-1} - \phi_1 \mu_{t-1})$. If ℓ is the identity link, then $\text{E}(Y_t \mid \boldsymbol{X}_t) = \mu_t$. However, no other link functions give this neat result.

It seems the only model of the form (3.7) which gives an interpretable marginal mean has ℓ equal to the identity function (as in Grunwald and Hyndman, 1998). But then there is always a danger of the mean lying outside $[0, 1]$.

Of course, it is possible to satisfactorily interpret the mean conditional on $\boldsymbol{D}_t$ when non-identity links functions are used, but this seems unnecessarily complicated and provides no advantage over the simpler models of the form (3.1).

3.3 Binary additive models with autocorrelated errors

A third proposal which seeks to allow autocorrelation error properties with the marginal interpretation of the model (3.6) is defined as follows. Let $Y_t \mid \boldsymbol{D}_t \sim \text{Bin}(m_t)$ where $m_t =$

$\min(\max(m_t^*, 0), 1)$,

$$m_t^* = \mu_t + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu_{t-j}) \qquad \text{and} \qquad \ell(\mu_t) = \lambda + \sum_{i=1}^{r} g_i(X_{i,t}) \tag{3.8}$$

where ℓ denotes the logistic link function.

When $r = 0$ is a constant, this model is equivalent to the binary AR(p) defined by (2.1).

Let $p_t = \Pr(0 < m_t^* < 1)$ be the probability that m_t^* takes permissible values. Then,

$$\mathrm{E}(Y_t) = p_t \left[\mu_t + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu_{t-j}) \right] + \Pr(m_t^* > 1).$$

So if $p_t \approx 1$, then $\mathrm{E}(Y_t) \approx \mu_t$ and each of the terms g_i can be interpreted as the effect of $X_{i,t}$ on the marginal mean of Y_t.

The use of the link function in the equation for μ_t does not ensure that m_t^* takes only permissible values for all possible values of $\boldsymbol{X}_t$. However, it increases the probability p_t of it doing so.

Note that when m_t^* is permissible, the model can also be expressed with $Y_t = \mu_t + e_t$, where

$$e_t = \phi_1 e_{t-1} + \cdots + \phi_p e_{t-p} + \delta_t, \tag{3.9}$$

and $\{\delta_t\}$ is an independent series with zero mean. Thus the errors from the model follow an autoregressive process. In particular, the Yule-Walker equations (2.5) will hold where $\gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h} \mid \boldsymbol{X}_t)$.

This model is related to the first order autocorrelated logistic regression model proposed by Zeger, Liang and Self (1985).

While this model appears to possess some useful properties, and retains those properties if it is generalized to other distributions, there is much work to be done in establishing estimation algorithms and inference before this model can be of use in applications. Consequently, I have not used it in the applications described in the following sections.

4 Modelling IBM share transactions

I shall illustrate the transitional binary additive model using transaction occurrence for the frequently traded IBM stock. The data were taken from Engle and Russell (1998). The data used here concern transactions on the consolidated market for 1 November 1990. Each transaction is recorded including bid and ask quote movements, the volume associated with the transactions and the transaction prices. The time measured in seconds after midnight is also recorded.

There were 757 trades at 690 unique times. In many cases, there are multiple buyers or sellers involved in a trade and trades can have a widely differing number of shares

transferred. On 1 November 1990, the number of shares transferred ranged from 100 to 38200 at each trade.

We aggregate the data to 60 second intervals and let Y_t denote the occurrence of at least one transaction during minute t, $t = 1, \ldots, 390$. Also let X_t denote the square root of the number of shares traded during minute t, let P_t denote the average IBM share price during minute t. Then we consider the model $Y_t|\boldsymbol{D}_{t-1} \sim \text{Bin}(m_t)$ where

$$\ell(m_t) = \lambda + \sum_{j=1}^{p} \phi_j Y_{t-j} + g_1(X_{t-1}) + g_2(P_{t-1}) + g_3(t),$$

ℓ denotes the logistic function (2.2) and each g_i ($i = 1, 2, 3$) is a smooth function.

I fitted this model for $0 \leq p \leq 5$ using the `gam` function in S-Plus with cubic smoothing splines to estimate each g_i. For these models, the AIC can be defined as

$$\text{AIC} = \text{deviance} + 2(p + 1 + \text{df}_1 + \text{df}_2 + \text{df}_3)$$

where df_i denotes the equivalent degrees of freedom for the smooth function g_i. Hastie and Tibshirani (1990) give details on how these can be computed. For this example, the smoothing parameter for the splines was selected so that the $\text{df}_i = 4$ for all i. The model with $p = 1$ minimized the AIC. For the fitted model, $g_3(t)$ and $g_2(P_{t-1})$ are not significant. So our revised model has

$$\ell(m_t) = \lambda + \phi_1 Y_{t-j} + g_1(X_{t-1}).$$

The fitted coefficients are $\hat{\lambda} = -1.60$ and $\hat{\phi}_1 = 0.35$.

So if there was no transaction in the previous period, the probability of a transaction in the current period is

$$\ell^{-1}\left(-1.60 + g_1(0)\right) = 0.168.$$

Over the entire day, 47.8% of minutes included at least one transaction.

If there was at least one transaction in the previous period, we can study the probability of transaction as a function of the number of shares traded in the previous period. Figure 1 shows this relationship with pointwise 95% confidence intervals (see Hastie and Tibshirani, 1990, for details on how these are computed.) Clearly, stock transaction occurrences are more likely if there have been large transactions in the previous time period.

This is only a preliminary analysis of what could be done with transaction data. For example, it is of interest to investigate the intra-day patterns, and the effect of the bid and ask prices. It would be possible to include the number of transactions in the previous 60 seconds rather than only the number of shares traded. It may also be preferable to make the covariates moving 60 second totals rather than enforce fixed intervals.

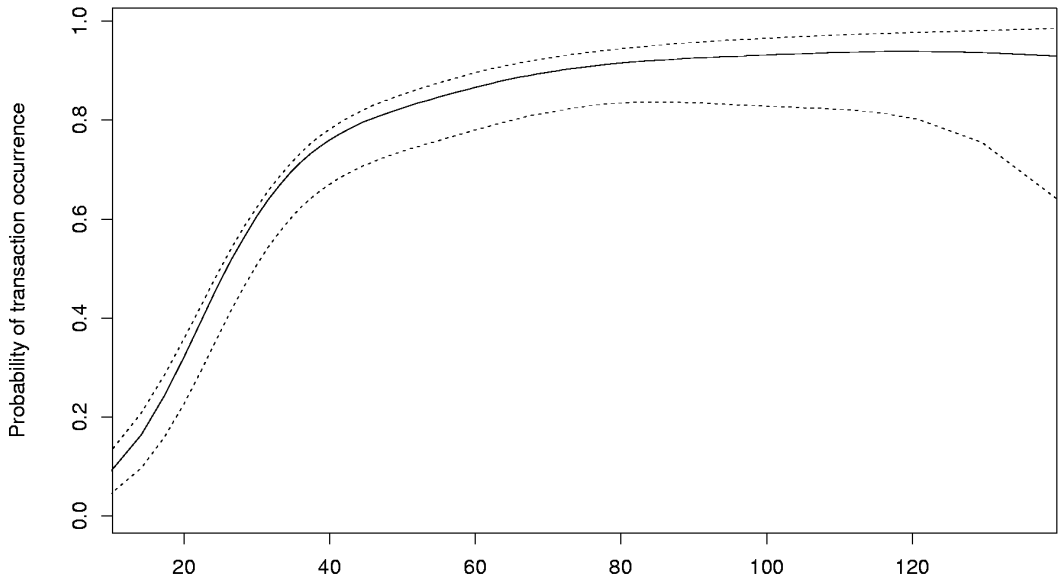

Figure 1: *The estimated probability of transaction following a minute in which a transaction occurred. The dashed lines represent pointwise 95% confidence intervals for the fitted curve.*

5 Modelling rainfall occurrence in Melbourne

As a further application of the models discussed here, I consider the daily rainfall data from Melbourne, Australia (the Melbourne city station, 86071) for the period 1 January 1963 to 30 September 1998. A more detailed analysis of these data is given in Hyndman and Grunwald (1998). During this period rainfall was recorded on 39.8% of the 13,057 days. I will show that for this series, rainfall occurrence is influenced by several factors including seasonality, drought, and rainfall occurrence and intensity the preceding day.

Stern and Coe (1984) model rainfall occurrence using a special case of the transitional binary additive model with all g_i linear. These methods are effective in describing typical rainfall patterns throughout the year, but they assume the same seasonal pattern for each year and thus are not capable of highlighting droughts, trends, or other effects not well-modelled by periodic seasonal patterns. They also do not identify interesting non-linearity in the relationships. Using an additive (and nonparametric) formulation, we can relax the assumption that each year follows the same seasonal pattern and we can allow non-linear relationships.

Let Q_t denote the quantity of rain on day t and let Y_t denote the occurrence of rain on day t ($Y_t = 1$ if $Q_t > 0$; otherwise $Y_t = 0$). To simplify the analysis of seasonality, I omitted the 9 leap days from the series, although the leap day data were used as the lagged regressors on March 1 when it followed a leap day. As in Grunwald and Jones (1998), I use a transformation of previous rainfall values to improve the fit. Specifically, let $W_t = \log(Q_t + c)$. (Without this transformation, a variable bandwidth would be necessary due to the extreme skewness of Q_t.) For the GLMs fitted by Grunwald and Jones, c was chosen by maximum likelihood to be equal to 0.2. To facilitate comparisons between models, I shall also use $c = 0.2$ in this paper.

Define the vector of covariates

$$\boldsymbol{D}_t = (Y_{t-1}, \ldots, Y_{t-p}, W_{t-1}, \ldots, W_{t-p}, X_{1,\,t-1}, \ldots, X_{r-p,\,t-1}, t)'$$

and let $\ell(m_t) = f_t(\boldsymbol{D}_{t-1})$ where

$$f_t(\boldsymbol{D}_{t-1}) = \lambda + \sum_{j=1}^{p} (\phi Y_{t-j} + g_j(W_{t-j})) + \sum_{i=p+1}^{r} g_i(X_{i-p,\,t-1}) + g_{r+1}(t),$$

ℓ denotes the logistic function (2.2) and each g_i ($i = 1, 2, \ldots, r+1$) is a smooth function.

I included the covariate $X_{1,\,t-1} = I_t$ where I_t denotes the value of the Southern Oscillation Index (SOI), the standardized anomaly of the Mean Sea Level Pressure (MSLP) between Tahiti and Darwin. If T_k denotes the Tahiti MSLP and D_k denotes the Darwin MSLP for month k, then the monthly value of I_k is calculated as $I_k = 10(T_k - D_k - \mu_k)/\sigma_k$ where μ_k denotes the long term average of $(T_k - D_k)$ for that month and σ_k denotes the standard deviation of $T_k - D_k$ for that month. (This is known as the Troup SOI.) Figure 2 shows the monthly values between January 1963 and September 1998. There is clearly a lot of random variation in the measurement. I have highlighted the underlying trend with a loess curve (Cleveland, Grosse and Shyu, 1992) of degree 2 and span 6%. Negative values of I_k indicate "El Niño" episodes and are usually accompanied by sustained warming of the central and eastern tropical Pacific Ocean, a decrease in the strength of the Pacific Trade Winds, and a reduction in rainfall over eastern and northern Australia. Positive values of I_k are associated with stronger Pacific trade winds and warmer sea temperatures to the north of Australia (a "La Niña" episode). Together these are thought to give a high probability that eastern and northern Australia will be wetter than normal. It should be noted that the effect of the Southern Oscillation is greater in Queensland and New South Wales than Victoria (Allan, Lindesay and Parker, 1996). Define I_t to be the value of the fitted loess curve at day t. (Almost identical results are obtained if I_t is calculated by linearly interpolating the raw values of I_k.)

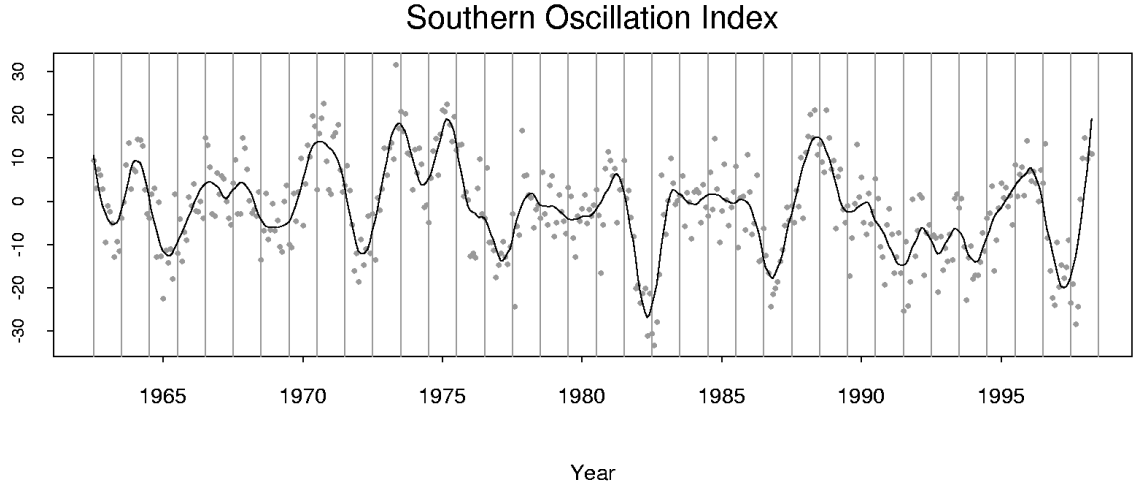

Figure 2: *Monthly Southern Oscillation Index with smooth line highlighting the pattern. The smooth line was computed using a loess curve of degree 2 with span of 6%.*

I also included the covariate $X_{2,\,t-1} = S_t = t \bmod 365$ to model the seasonal variation. The function g_{p+2} is constrained to be periodic; that is, we constrain $g_{p+2}(S_t)$ to be smooth at the boundary between $S_t = 365$ and $S_t = 1$.

Thus our model has

$$f_t(\boldsymbol{x}_{t-1}) = \lambda + \sum_{j=1}^{p}(\phi_j Y_{t-j} + g_j(W_{t-j}) + g_{p+1}(I_t) + g_{p+2}(S_t) + g_{p+3}(t).$$

Models with $p = 1, 2, 3$ and 4 were fitted. The results were very similar for all p so I selected $p = 1$ as it simplifies the interpretation.

The smooth term involving I_t was not significantly different from a linear function and so $g_2(I_t)$ was restricted to the linear function $g_2(I_2) = \beta I_2$. Because $g_3(S_t)$ is a periodic function, we model it using a Fourier function of the form

$$g_3(S_t) = \sum_{k=1}^{m} \left[\beta_{i,s} \sin(2\pi k S_t/365) + \beta_{i,c} \cos(2\pi k S_t/365)\right],$$

and select the value of m using the AIC. (An alternative approach would be to use a periodic smoother.) The smooth terms $g_1(W_{t-1})$ and $g_4(t)$ were fitted using smoothing splines. The final model had $\phi_1 = 0.26$, $\beta = 0.0088$, $m = 3$ in $g_3(S_t)$ and smoothing parameters $\text{df}_1 = 4.9$ and $\text{df}_4 = 50$ where df_i denotes the degrees of freedom for the smooth function g_i. The value of df_1 was chosen by minimizing the AIC, while the smoothing parameter for $g_4(t)$ was selected to allow sufficient flexibility to model changes in the probability of occurrence over a period of two or three years.

The value of β was significant (using a t-test at the 5% level). However, if the SOI term was omitted from the model and the other terms re-estimated, the deviance of the model did not change significantly (using a χ^2 test at the 5% level). This anomaly occurs because, if SOI is omitted, the $g_4(t)$ term can model the variation in SOI. We choose to include SOI because we are interested in assessing its effect on rainfall.

Figure 3 shows some results for the fitted model. The lower solid line is the estimate of the probability of rain following a dry day ($y_{t-1} = 0$):

$$\Pr(Y_t = 1 \,|\, Q_{t-1} = 0) = \ell(\lambda + g_1(\log c) + g_2(I_t) + g_3(S_t) + g_4(t)).$$

The upper solid line is the estimate of the probability of rain following a day of median intensity (2mm):

$$\Pr(Y_t = 1 \,|\, Q_{t-1} = 2) = \ell(\lambda + \phi_1 + g_1(\log(2 + c)) + g_2(I_t) + g_3(S_t) + g_4(t)).$$

For comparison, analogous curves for a GLM are shown as dashed lines. This model had

$$f_t(\boldsymbol{D}_{t-1}) = \lambda + \phi_1 Y_{t-1} + \phi_1^* W_{t-1} + \beta I_t + \sum_{k=1}^{3} \left[\beta_{i,s} \sin(2\pi k S_t/365) + \beta_{i,c} \cos(2\pi k S_t/365)\right].$$

Again, the AIC was used to select the number of sinusoidal terms in the seasonal pattern.

Higher order AR models were tried but gave very similar results. Note that the GAM allows the modelling of non-seasonal temporal variation whereas the GLM does not.

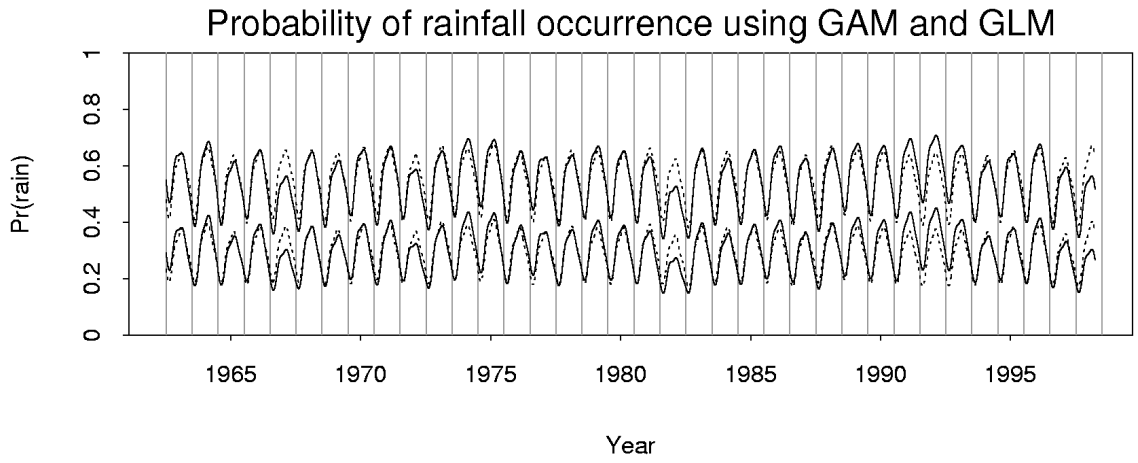

Figure 3: *Lower solid line: estimated probability of rain following a dry day. Upper solid line: estimated probability of rain following a day of median intensity (2mm). These estimates are based on the GAM; dashed lines show analogous curves for the GLM.*

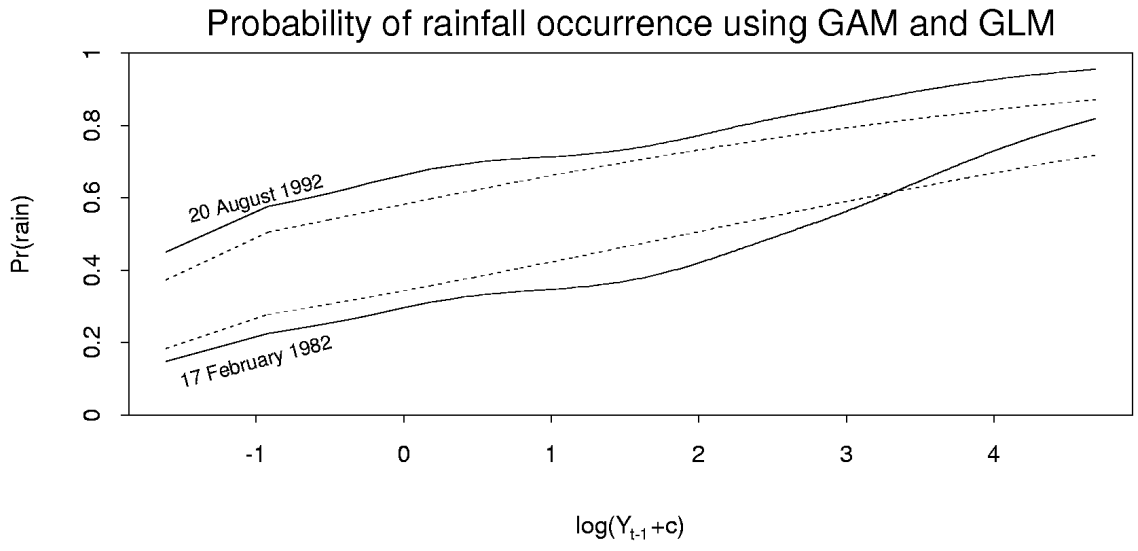

Figure 4: *Lower solid line: estimated probability of rain on 17 February 1982. Upper solid line: estimated probability of rain on 20 August 1992. Dashed lines show analogous curves for the GLM.*

We can also look at the probability of rainfall occurrence as a function of the rainfall intensity of the previous day. Figure 4 shows this relationship for two days in the period of the data. The lower curves are for 17 February 1982 (when $g_2(I_t) + g_3(S_t) + g_4(t)$ was minimized). The upper curves are for 20 August 1992 (when $g_2(I_t) + g_3(S_t) + g_4(t)$ was maximized). The solid lines represent the probabilities calculated using the GAM, conditioning on the value of t. The dashed lines show the analogous probabilities as calculated using the GLM.

5.1 Seasonally adjusted probabilities

One object of this analysis is to highlight unusual periods of occurrence, relative to "typical" annual occurrence patterns. For instance, comparing the GLM and GAM fits in Figure 3 suggests that 1982 had unusually low occurrence and 1990–1993 had unusually high occurrence. To facilitate and quantify such comparisons, we can apply a simple method of seasonal decomposition to decompose $f_t(\boldsymbol{D}_{t-1})$ into a seasonal term $s_t(\boldsymbol{D}_{t-1})$ that repeats each year and represents a "typical" year, and a remainder term $r_t(\boldsymbol{D}_{t-1})$ that represents deviations from this regular pattern. Let $f_t(\boldsymbol{D}_{t-1}) = s_t(\boldsymbol{D}_{t-1}) + r_t(\boldsymbol{D}_{t-1})$ where $s_t(\boldsymbol{D}_{t-1}) = s_{t+365k}(\boldsymbol{D}_{t-1})$ for $k = 1, 2, \ldots$. These effects can be interpreted in terms of odds of rain, so that

$$\frac{\Pr[Y_t = 1 \,|\, \boldsymbol{D}_{t-1}]}{\Pr[Y_t = 0 \,|\, \boldsymbol{D}_{t-1}]} = \exp\{f_t(\boldsymbol{D}_{t-1})\} = \exp\{s_t(\boldsymbol{D}_{t-1})\} \exp\{r_t(\boldsymbol{D}_{t-1})\} \,.$$

Thus $\exp\{r_t(\boldsymbol{D}_{t-1})\}$ represents the factor deviation of the odds of rain from the odds in a typical year, at time t. The seasonally adjusted probability of rain is

$$\pi_t^a(\boldsymbol{D}_{t-1}) = \ell(\bar{s}(\boldsymbol{D}_{t-1}) + r_t(\boldsymbol{D}_{t-1}))$$

where $\bar{s}(\boldsymbol{D}_{t-1}) = \frac{1}{365} \sum_{t=1}^{365} s_t(\boldsymbol{D}_{t-1})$, and the seasonal probability of rain is

$$\pi_t^s(\boldsymbol{D}_{t-1}) = \ell(s_t(\boldsymbol{D}_{t-1})) \,.$$

Our model provides a convenient estimate of $s_t(\boldsymbol{D}_{t-1})$. We let

$$\hat{s}_t(\boldsymbol{D}_{t-1}) = \hat{\lambda} + \hat{\phi}_1 \bar{Y}_{t-1} + \bar{g}_1 + \hat{\beta}\bar{I} + \hat{g}_3(S_t) + \bar{g}_4$$

where $\bar{Y}$ denotes the mean of Y_t, $\bar{I}$ denotes the mean of I_t, $\bar{g}_1$ denotes the mean of $\hat{g}_1(W_{t-1})$ and $\bar{g}_4$ denotes the mean of $\hat{g}_4(t)$, $t = 1, \ldots, n$.

Figure 5 shows estimates of the seasonal probability of rain, π_t^s, and the seasonally adjusted probability of rain, π_t^a, plotted against time t. The curves for $Q_{t-1} = 0$ and $Q_{t-1} = 2$mm are shown. The most striking periods of low occurrence are in 1967, 1972, 1982 and 1998. Apart from the most recent drought, these are exactly the droughts in areas encompassing Melbourne, as reported by Keating (1992). The period of highest probability of occurrence is 1992 (which had the greatest number of wet days of any year in the period studied).

Our model attempts to separate the non-seasonal temporal variation, $g_2(I_t) + g_4(t)$, into two parts: one due to the SOI and one which is unaffected by the SOI. To help visualize

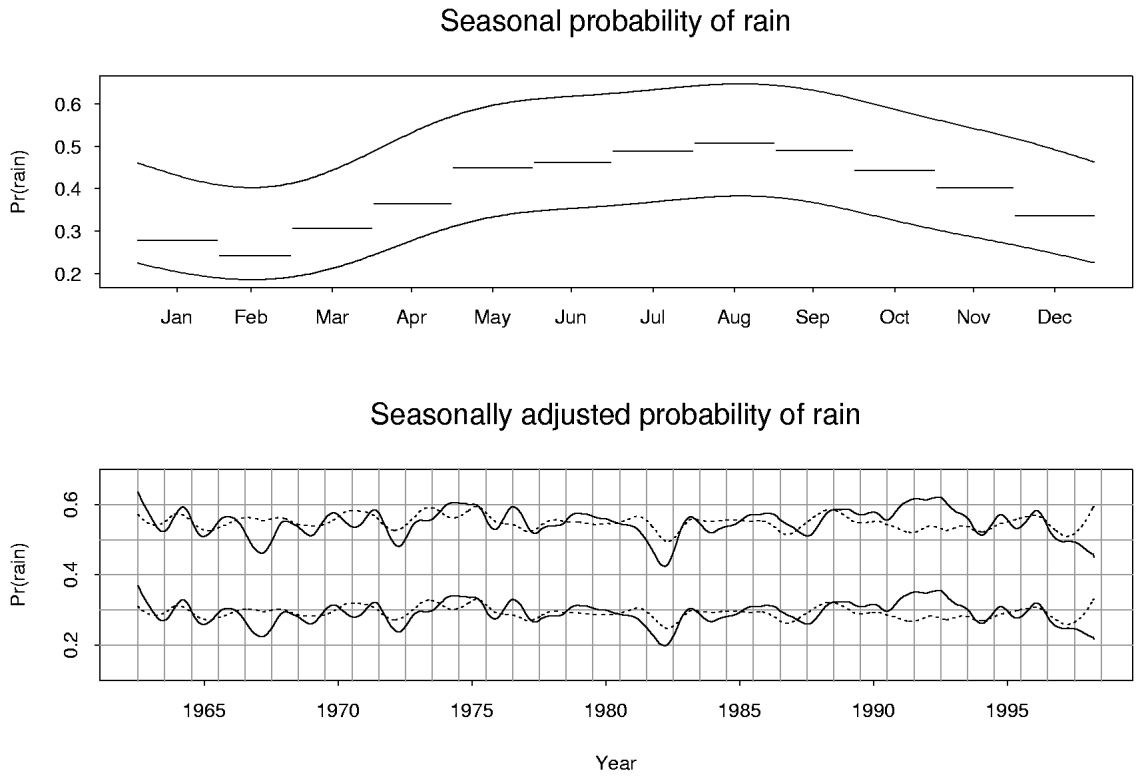

Figure 5: *Top: Estimated seasonal probability of rain, π_t^s. The horizontal bars show the proportion of rainy days for each month during the data period. Bottom: seasonally adjusted estimated probability of rain, π_t^a. Upper solid lines show curves following a dry day; lower solid lines show curves following a day of median intensity (2mm). The dashed lines shows the estimated probability of rain further adjusted to show the effect of the SOI.*

the effect of this separation, the dashed lines in the bottom plot of Figure 5 show the probability of rain predicted by the model after seasonal adjustment and removing the effect of $g_4(t)$. That is, we plot the estimate of

$$\ell(\lambda + \phi_1 Y_{t-1} + g_1(W_{t-1}) + \beta I_2 + \bar{g}_3 + \bar{g}_4).$$

The resulting curve shows the effect of the SOI on rainfall probability.

The differences between the solid and dashed curves are of interest. For example, in 1967, the solid curve is substantially lower than the dashed curve. This was a period of drought (reflected by the dip in the solid curve) which was not associated with a corresponding low in SOI. The drought of 1982 was associated with the SOI (hence the trough in the dashed curve), but it was more severe than the SOI suggested. Thus, the solid line dips further than the dashed line. The period 1991–1993 is one with unusually high rainfall occurrence that was not associated with a corresponding high in the SOI.

Much of the non-seasonal temporal variation in rainfall probability is being modelled by $g_4(t)$ rather than $g_2(I_t)$. So while the SOI appears to have some effect on the rainfall occurrence it is not a strong predictor and extreme values of the SOI do not always translate into extreme values of rainfall probability.

Currently, the Bureau of Meteorology uses forecasts of the SOI to guide their long-term climate prediction. The analysis presented here suggests that that procedure is not going to yield good prediction for southern Victoria because the relationship between SOI and rainfall occurrence is not strong. The method is probably much better for locations in New South Wales and Queensland where the relationship between SOI and rainfall is stronger (Allan, Lindesay and Parker, 1996). However, even there the model presented here will probably lead to better long-term forecasts as it incorporates temporal variation not due to the SOI.

6 Conclusions

I have introduced autoregression and nonparametric additive regression models for binary time series. For the binary AR(p) model, I have provided stationarity conditions and Yule-Walker equations, and demonstrated the equivalence between the binary AR(1) model and several previously proposed models. Various possible models for nonparametric additive regression with autocorrelation have been considered, and it has been shown that simply including lagged values of the response variable in a logistic additive regression provides interpretable results. Furthermore, extensive inferential methods developed for generalized additive modelling are readily applicable.

7 Acknowledgements

This work was supported by Australian Research Council grants. The work on modelling rainfall was done while the author was visiting the Department of Statistics, Colorado State University.

8 References

ALLAN, R., LINDESAY, J. and PARKER, D. (1996) *El Niño: Southern oscillation and climatic variability*. CSIRO Publications, Melbourne, Australia.

ALTMAN, N.S. (1990) Kernel smoothing of data with correlated errors, *J. Amer. Statist. Assoc.*, **85**, 749–759.

CLEVELAND, W.S., GROSSE, E. and SHYU, W.M. (1992) Local regression models, in J.M. Chambers and T.J. Hastie (eds.), *Statistical models in S*, Wadsworth and Brooks: Pacific Grove.

COX, D.R., and SNELL, E.J. (1989) *Analysis of binary data*, 2nd ed., Chapman and Hall: London.

DIGGLE, P.J., and HUTCHINSON, M.F. (1989) On spline smoothing with autocorrelated errors, *Aust. J. Statistics*, **31**, 166–182.

DIGGLE, P.J., LIANG, K.L., and ZEGER, S.L. (1994) *Analysis of longitudinal data*, Oxford University Press.

EILERS, P.H.C. and MARX, B.D. (1996) Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.*, **89**, 89–121.

ENGLE, R.F., and RUSSELL, J.R. (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica*, **66** (5), 1127–1162.

FAN and GIJBELS (1996) *Local polynomial modelling and its applications*, Chapman and Hall: London.

GREEN, P.J. and SILVERMAN, B.W. (1994) Nonparametric regression and generalized linear models: a roughness penalty approach, Chapman and Hall: London.

GRUNWALD, G.K., and HYNDMAN, R.J. (1998) Smoothing non-Gaussian time series with autoregressive structure, *Computational Statistics and Data Analysis*, **28**, 171–191.

GRUNWALD, G.K., HYNDMAN, R.J., TEDESCO, L., and TWEEDIE, R.L. (1997) A unified view of linear AR(1) models, Statistics research report, Department of Mathematics and Statistics, Monash University. Available on the web at: `www-personal.buseco.monash.edu.au/~hyndman/papers`.

GRUNWALD, G.K. and JONES, R.J. (1998) Markov models for time series with mixed distribution. *Environmetrics*, to appear.

HART, J.D. (1991) Kernel regression estimation with time series errors, *J. R. Statist. Soc. B*, **53**, 173–187.

HASTIE, T.J., and TIBSHIRANI, R.J. (1990) *Generalized Additive Models*, Chapman and Hall: London.

HYNDMAN, R.J., and GRUNWALD, G.K. (1998) Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall, Working paper, Department of Econometrics and Business Statistics, Monash University. Available on the web at: `www-personal.buseco.monash.edu.au/~hyndman/papers`.

JACOBS, P.A. and LEWIS, P.A.W. (1978a) Discrete time series generated by mixtures I: correlational and runs properties, *J. Roy. Statist. Soc. Ser. B*, **40**, 94–105.

JACOBS, P.A. and LEWIS, P.A.W. (1978b) Discrete time series generated by mixtures II: asymptotic properties, *J. Roy. Statist. Soc. Ser. B*, **40**, 222–228.

KANTER, M. (1975) Autoregression for discrete processes mod 2, *J. Appl. Prob.*, **12**, 371–375.

KEATING, J. (1992) *The drought walked through: a history of water shortage in Victoria*. Department of Water Resources Victoria: Melbourne.

KEDEM, B. (1980a) *Binary time series*, Marcel Dekker: New York.

KEDEM, B. (1980b) Estimation of the parameters in stationary autoregressive process after hard limiting, *J. Amer. Statist. Assoc.*, **75**, 146–153.

KEENAN, D.M. (1982) A time series analysis of binary data, *J. Amer. Statist. Assoc.*, **77**, 816–820.

LEWIS, P.A.W. (1980) Simple models for positive-valued and discrete-valued time series with ARMA correlation structure, In *Multivariate analysis V*, ed. P.R. Krishnaiah, pp.151–166, North Holland.

LI, W.K. (1991) Testing model adequacy for some Markov regression models for time series, *Biometrika*, **71**, 83–89.

LI, W.K. (1994) Time series models based on generalized linear models: some further results, *Biometrics*, **50**, 506–511.

LOMNICKI, Z.A., and ZAREMBA, S.K. (1955) Some applications of zero-one processes, *J. Roy. Statist. Soc., Ser. B*, **17**, 243–255.

MCCULLAGH, P., and J.A. NELDER (1989) *Generalized Linear Models*, 2nd ed., Chapman and Hall: London.

MCKENZIE, E. (1985) Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.

NIU, X.F. (1996) Nonlinear additive models for environmental time series, with applications to ground-level ozone data analysis, *J. Amer. Statist. Assoc.*, **91**, 1310–1321.

STERN, R.D. and COE, R. (1984) A model fitting analysis of daily rainfall data (with discussion), *J. R. Statist. Soc. A*, **147**, 1–34.

ZEGER, S.L., LIANG, K.-Y., and SELF, S.G. (1985) The analysis of binary longitudinal data with time-independent covariates, *Biometrika*, **72**, 31–38.

ZEGER, S.L., and QAQISH, B. (1988) Markov regression models for time series: a quasi-likelihood approach, *Biometrics*, **44**, 1019–1031.