Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

The analysis of ordinal time-series data via a transition (Markov) model

Kathryn Bartimote-Aufflick [a] & Peter C. Thomson [b]

[a] Institute for Teaching and Learning, Carslaw F07 , University of Sydney , NSW, 2006, Australia

[b] Faculty of Veterinary Science , University of Sydney , Private Bag 4003, Narellan, NSW, 2567, Australia

Published online: 20 Nov 2010.

To cite this article: Kathryn Bartimote-Aufflick & Peter C. Thomson (2011) The analysis of ordinal time-series data via a transition (Markov) model, Journal of Applied Statistics, 38:9, 1883-1897, DOI: 10.1080/02664763.2010.529885

To link to this article: http://dx.doi.org/10.1080/02664763.2010.529885

Journal of Applied Statistics
Vol. 38, No. 9, September 2011, 1883–1897

Taylor & Francis
Taylor & Francis Group

The analysis of ordinal time-series data via a transition (Markov) model

Kathryn Bartimote-Aufflick[a]* and Peter C. Thomson[b]

[a]*Institute for Teaching and Learning, Carslaw F07, University of Sydney NSW 2006, Australia;* [b]*Faculty of Veterinary Science, University of Sydney, Private Bag 4003, Narellan NSW 2567, Australia*

(*Received 13 October 2009; final version received 8 September 2010*)

While standard techniques are available for the analysis of time-series (longitudinal) data, and for ordinal (rating) data, not much is available for the combination of the two, at least in a readily-usable form. However, this data type is common place in the natural and health sciences where repeated ratings are recorded on the same subject. To analyse these data, this paper considers a transition (Markov) model where the rating of a subject at one time depends explicitly on the observed rating at the previous point of time by incorporating the previous rating as a predictor variable. Complications arise with adequate handling of data at the first observation ($t = 1$), as there is no prior observation to use as a predictor. To overcome this, it is postulated the existence of a rating at time $t = 0$; however it is treated as 'missing data' and the expectation–maximisation algorithm used to accommodate this. The particular benefits of this method are shown for shorter time series.

Keywords: E–M algorithm; ordinal; repeated measures; transition model; longitudinal; categorical

1. Introduction

Ratings or scores are an example of what are known as 'ordinal' data or equivalently 'ordered categorical' data. In this paper, we consider ratings that are made repeatedly on the same subject (or experimental unit). Time-series (or 'longitudinal', 'serial', 'panel') data provide an additional complication to an analysis, since we need to account for the correlation between measurements (at different times) on the same subject.

Methodologies for analysing either ordinal data, or time-series data, are well documented. For ordinal data, the technique of choice is most often ordinal logistic regression (OLR) [1,7,13]. For normally distributed time-series data, restricted maximum-likelihood estimation [14,19] is often used (incorporating a model for correlation within subject across time) in the pursuit of less-biased estimates of the variance. Less well known are methodologies for dealing with both complexities (ordinal data and time series or repeated measurements) simultaneously.

*Corresponding author. Email: kathryn.aufflick@sydney.edu.au

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664763.2010.529885
http://www.informaworld.com

To analyse discrete time-series data, such as ordinal time-series data, we move towards extensions of generalised linear models (GLMs) [13]. The three main approaches used to model discrete time-series data are: marginal models; random effects models; and transition (or Markov) models, the latter where the dependence on prior observations is explicitly modelled. It is important to note that for linear models, the interpretation of the regression parameters is independent of the correlation structure. However, with GLMs for discrete data, different assumptions about the source of correlation (and hence the chosen correlation model) can lead to regression coefficients for the same data set with different values and interpretations [4]. Therefore, it is important to give careful thought to the source of correlation in one's data and the level at which you wish to be able to interpret. Diggle *et al*. [4] outline these three approaches in the context of binary and count data.

We have chosen to use a transition (Markov) model (using OLR) as described in Diggle *et al*. [4] for the analysis of the example data (disease ratings in cocoa trees) for the following reasons: (1) For this particular data set, it is biologically intuitive that disease rating (in the current time) depends explicitly on the status of the tree in the previous time; (2) This provides a conceptually simple method that models the prior dependence in the data structure (as opposed to modelling the correlation structure) and can be readily understood by agricultural scientists and other non-statisticians. It is straightforward for non-statisticians to implement in its basic form [26], however if a more complex approach is chosen (the work described in this paper) information from the first observation can be incorporated to include the full likelihood using a parametric approach; (3) The analysis can be achieved in existing software – we have used the public domain software R [20].

Some additional treatments of transition models for time-series ordinal data can be found in several papers [9,11,25]. However, unlike those papers, the method being outlined here fully accommodates information from the first observation, an aspect not covered when initially described by Diggle *et al*. [4]. It should be noted that a generalised estimating equation (GEE) approach [6,10,16,17], an implementation of the marginal model, could also be used to model ordinal longitudinal data, as has been illustrated, for example, with plant disease and horticultural applications [2,18] and in gene mapping applications [8,24]. However, while widely used in medical statistics, this approach is not so commonly used in agricultural statistics which is the focus of the application in this paper. For example, data may frequently be analysed using non-parametric methods in the agricultural sciences [22].

Section 2 of this paper outlines how the ordinal logistic model can be extended to a transition model, and then considers various approaches for fitting such models. This includes conditioning on the first observation (Section 2.2) and then develops the methods for a full likelihood calculation which treats unrecorded ratings prior to the first observation as a missing data process. A model-based method to calculate this distribution (at time $t = 0$) is considered in Section 2.3. Section 3 then provides an illustration of the use of this technique, taken from a horticultural application. Some general issues related to these techniques are then discussed in Section 4.

2. Development of the ordinal transition model

We will assume that each observation depends on the previous observation, but not on any earlier observation. To introduce some notation (following that of Diggle *et al*. [4]), we let Y_{it} represent a rating response variable for subject i at time t. The rating categories take the values $k = 0, \ldots, C-1$ (with a total of C categories). For the transition model, the conditional distribution of Y_{it}, given $Y_{i,t-1} = y_{i,t-1}$ is

$$\text{logit } P(Y_{it} \leq k | Y_{i,t-1} = y_{i,t-1}) = \theta_k - (\boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\alpha}' \mathbf{y}^*_{i,t-1}), \tag{1}$$

for $k = 0, 1, \ldots, C-2$; $i = 1, 2, \ldots, n$; $t = 2, 3, \ldots, T$ where θ_k is the intercept relating to the kth ordered rating where $\theta_0 \leq \theta_1 \leq \theta_2 \cdots \leq \theta_{C-2}$, $\boldsymbol{\beta}$ is the vector of regression coefficients corresponding to the set of predictor variables, $\mathbf{x}_{it}$, and $\boldsymbol{\alpha}$ is a vector of regression coefficients reflecting the magnitude of dependence on the previous observation. This model is reduced compared with that outlined by Diggle *et al*. [4, Section 10.3.2] who allow the effect of the covariates to change depending on the value of the previous observation; however, the method outlined here to easily be expanded to accommodate fitting of the full model. Note the negative sign included for the regression coefficients implies that a positive regression coefficient β_j is associated with an increase in the rating on average, for a unit increase in x_j. The model is frequently written without this convention and both forms are seen in software implementation of ordinal logistic models. It should also be noted that the commonly used proportional odds model has been adopted here, the assumption being that the cumulative odds for events $Y \leq k$ and $Y \leq k'$ for the same values $\mathbf{x}_{it}$ and $\mathbf{y}^*_{i,t-1}$ are in proportion, i.e. $\exp(\theta_k/\theta_{k'})$. However, other methods for handling ordinal data [1] could be incorporated in the current framework.

Here, $\mathbf{y}^*_{i,t-1} = \left(y^*_{i,t-1,0}, y^*_{i,t-1,1}, \ldots, y^*_{i,t-1,C-2}\right)'$ consists of a set of 0–1 indicator variables for the previous observation $y_{i,t-1}$ defined as

$$y^*_{i,t-1,k} = \begin{cases} 1 & \text{if } y_{i,t-1} \leq k \\ 0 & \text{otherwise.} \end{cases}$$

For example, if $y_{i,t-1} = 1$ with $C = 5$ ordered ratings, then $\mathbf{y}^*_{i,t-1} = (0, 1, 1, 1)'$.

Note that the model above, Equation (1), is specified as the cumulative probability of obtaining a certain rating or less. However, in order to obtain the event probability of obtaining a particular rating, e.g. $P(Y = 3)$, this is calculated as

$$\begin{aligned} &P(Y_{it} = k | Y_{i,t-1} = y_{i,t-1}) \\ &\quad = \begin{cases} P(Y_{it} = 0 | Y_{i,t-1} = y_{i,t-1}) & k = 0 \\ P(Y_{it} \leq k | Y_{i,t-1} = y_{i,t-1}) - P(Y_{it} \leq k-1 | Y_{i,t-1} = y_{i,t-1}) & k = 1, \ldots, C-2 \\ 1 - P(Y_{it} \leq C-2 | Y_{i,t-1} = y_{i,t-1}) & k = C-1 \end{cases} \end{aligned}$$

2.1 *Methods of fitting the ordinal transition model*

The complexity lies not in the description of the model, but in the fitting of it. We would like to obtain the full likelihood

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{n} p(y_{i1}, y_{i2}, \ldots, y_{iT}),$$

where $\boldsymbol{\Omega} = (\boldsymbol{\theta}', \boldsymbol{\beta}', \boldsymbol{\alpha}')'$ and $p(y_{i1}, y_{i2}, \ldots, y_{iT})$ is the joint probability distribution for the repeated ratings for a particular subject, i.e. $P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iT} = y_{iT})$. If we apply a first-order Markov model, we are able to obtain

$$\begin{aligned} p(y_{i1}, y_{i2}, \ldots, y_{iT}) &= p(y_{i1}) p(y_{i2}, \ldots, y_{iT} | y_{i1}) \\ &= p(y_{i1}) \times p(y_{i2} | y_{i1}) \times \ldots \times p(y_{iT} | y_{i,T-1}). \end{aligned}$$

However, the problem is that we do not have the marginal distribution at time $t = 1$, $p(y_{i1})$. The main aim of this paper is to present some possible solutions to this problem. The two main strategies presented are (1) ignore $p(y_{i1})$ and effectively condition on y_{i1}; and (2) select a model for $p(y_{i1})$.

In the first strategy, where we ignore $p(y_{i1})$, a simple lag model is employed. In this case, the response data lagged by one time point becomes a predictor variable in the model, using data at times $t = 2, \ldots, T$ only. This is effectively conditioning on y_{i1}.

In the second strategy, we create a model for $p(y_{i1}|y_{i0})$ with y_{i0} a missing data problem. That is, in utilising a lag (or transition) model, the response data from the first time point ($t = 1$) has no corresponding lagged rating (y_{i0}) and hence the corresponding predictor indicator variables ($\mathbf{y}^*_{i0}$) are set as missing values. Heuristically, this is appealing as all the information of the experiment, including at time 1, will be utilised. We will focus the majority of the remainder of this paper on exploring and extending methodology for this estimation. However, firstly we will show how to fit a simple transition model by conditioning on the first observation.

2.2 *Fitting the model by conditioning on the first observation*

Let the conditional likelihood be

$$L(\mathbf{\Omega}) = \prod_{i=1}^{n} p(y_{i2}, y_{i3}, \ldots, y_{iT}|y_{i1}),$$

where $p(y_{i2}, y_{i3}, \ldots, y_{iT}|y_{i1}) = P(Y_{i2} = y_{i2}, Y_{i3} = y_{i3}, \ldots, Y_{iT} = y_{iT}|Y_{i1} = y_{i1})$. This conditional probability function for subject i can be rewritten as

$$p(y_{i2}, y_{i3}, \ldots, y_{iT}|y_{i1}) = \prod_{t=2}^{T} p(y_{it}|y_{i,t-1}),$$

since being a first-order Markov chain, the dependence only extends to the previous observation. So the likelihood is calculated as

$$L(\mathbf{\Omega}) = \prod_{i=1}^{n} p(y_{i2}, y_{i3}, \ldots, y_{iT}|y_{i1}) = \prod_{i=1}^{n}\prod_{t=2}^{T} p(y_{it}|y_{i,t-1}).$$

It is now clear that this conditional likelihood can be maximised using standard methods, since it is expressed as the product of conditionally independent terms. In practice, it will be the log-likelihood that is maximised, i.e.

$$\log L(\mathbf{\Omega}) = \sum_{i=1}^{n}\sum_{t=2}^{T} \log p(y_{it}|y_{i,t-1}). \tag{2}$$

We have used the `polr()` (proportional odds logistic regression) function from the MASS library in R [20] to fit the model. This method will be subsequently labelled 'Method 0'.

2.3 *Fitting the model using the first observation, with y_{i0} a missing data problem*

When we condition on y_{i1} (as outlined in the previous section), there is some loss of information, as the ratings from $t = 1$ are not explicitly included as responses in the model. We would prefer to work with the full log likelihood,

$$\log L(\mathbf{\Omega}) = \sum_{i=1}^{n} \log p(y_{i1}) + \sum_{i=1}^{n}\sum_{t=2}^{T} \log p(y_{it}|y_{i,t-1})$$

since

$$p(y_{i1}, y_{i2}, \ldots, y_{iT}) = p(y_{i1}) \times p(y_{i2}, \ldots, y_{iT}|y_{i1}),$$

where the second term on the right-hand side of this expression is the contribution of the ith subject to the conditional likelihood expressed in Equation (2). The difficulty is that the transition model only provides the means of calculating the conditional terms, $p(y_{it}|y_{i,t-1})$, and not the marginal term $p(y_{i1})$. However, from the total probability law we can say that

$$p(y_{i1}) = \sum_{k=0}^{C-1} \pi_{ik}^{(0)} P(Y_{i1} = y_{i1}|Y_{i0} = k),$$

where

$$\pi_{ik}^{(0)} = P(Y_{i0} = k).$$

That is, we can model the marginal probability, $p(y_{i1})$, by summing over C possible component distributions for time $t = 0$ with corresponding mixing weights $\pi_{i0}^{(0)}, \pi_{i1}^{(0)}, \ldots, \pi_{i,C-1}^{(0)}$, where $\sum_{k=0}^{C-1} \pi_{ik}^{(0)} = 1$, i.e. a mixture distribution. An interpretation of the above is that the status of the subject prior to time $t = 1$ is unknown (for instance, disease may be present or not). However, there are experimental situations where a disease, for example, is introduced to non-diseased subjects at time 1 or immediately prior to the beginning of a trial. The transition model structure can be used for the C components $p(y_{i1}|y_{i0} = k)$, $k = 0, \ldots, C - 1$. So the full likelihood can now be written as

$$\log L(\boldsymbol{\Omega}) = \sum_{i=1}^{n} \log \sum_{k=0}^{C-1} \pi_{ik}^{(0)} p(y_{i1}|y_{i0} = k) + \sum_{i=1}^{n} \sum_{t=2}^{T} \log p(y_{it}|y_{i,t-1}). \tag{3}$$

Ignoring temporarily the calculation of the $\pi_{ik}^{(0)}$, standard OLR methods cannot be applied directly to the mixture component (i.e. the first part) of this log likelihood. The following section outlines the implementation of the expectation–maximisation (E–M) algorithm to facilitate model fitting using standard OLR software.

2.4 *Implementation using the E–M algorithm*

Mixture models can be fitted using standard procedures via the E–M algorithm [3,15]. This is a standard 'missing' or 'incomplete' data handling procedure, and is applicable since the y_{i0}, $i = 1, \ldots, n$ are not observed. It can be shown (Appendix A.1) that the estimating equations for $\hat{\boldsymbol{\Omega}}$ based on Equation (3) can be written as

$$\sum_{i=1}^{n} \sum_{k=0}^{C-1} \tau_{ik}^{(0)} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{i1}|y_{i0} = k) + \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{it}|y_{i,t-1}) = \mathbf{0}, \tag{4}$$

where

$$\tau_{ik}^{(0)} = \frac{\pi_{ik}^{(0)} p(y_{i1}|y_{i0} = k)}{\sum_{k'=0}^{C-1} \pi_{ik'}^{(0)} p(y_{i1}|y_{i0} = k')}$$

is the posterior probability of $Y_{i0} = k$, given the observed value of y_{i1}. From this, it is apparent that the estimating equations are in a standard (i.e. non-mixture) form, where weights $\tau_{ik}^{(0)}$ can be applied to the first part of Equation (4), corresponding to the C possible values of y_{i0}, i.e. $0, 1, \ldots, C - 1$. Note that the second part of Equation (4) is just the usual unweighted form of the likelihood estimating equations.

However, in order to compute the weights, $\tau_{ik}^{(0)}$, we need to make some assumptions about the form of the $\pi_{ik}^{(0)}$. There are two possible approaches that could be used, namely, empirical observed proportions (and variations of choosing which data subsets to base these on); and model-based methods. Some possible empirical options are shown in Appendix A.2 (labelled Methods 1–4). However, all these empirical approaches have limitations – they are either based on relatively few values such as subject-specific ones or, they may not be very reliable estimates for the particular subject at that particular time because they were based on aggregated data across all times. A model-based approach for estimating $\pi_{ik}^{(0)}$ is now considered to overcome these limitations.

2.5 *Model-based method for estimating* $\boldsymbol{\pi_{ik}^{(0)}}$

For the E–M algorithm, we need to specify the marginal probability distribution for the unobserved ratings at time $t = 0$ for subject i, namely,

$$\pi_{ik}^{(0)} = P(Y_{i0} = k), \quad k = 0, 1, \ldots, C-1.$$

The method to be outlined here is a model-based one, where the fitted OLR model is used to construct the probability distribution. As a working model, we will assume that the marginal probability distribution is stationary at time $t = 0$, meaning that it is the same at the prior time(s), say at time $t = -1$. Using the same procedures for evaluating $\pi_{ik}^{(1)}$, we have

$$\begin{aligned}\pi_{ik}^{(0)} &= \sum_{j=0}^{C-1} P(Y_{i,-1} = j, Y_{i0} = k) \\ &= \sum_{j=0}^{C-1} P(Y_{i,-1} = j) P(Y_{i0} = k | Y_{i,-1} = j) \\ &= \sum_{j=0}^{C-1} \pi_{ij}^{(-1)} P(Y_{i0} = k | Y_{i,-1} = j).\end{aligned}$$

Under the stationarity assumption, we have $\pi_{ik}^{(0)} = \pi_{ik}^{(-1)}$, $j = 0, 1, \ldots, C-1$. So

$$\pi_{ik}^{(0)} = \sum_{j=0}^{C-1} \pi_{ij}^{(0)} P(Y_{i0} = k | Y_{i,-1} = j).$$

If we consider the set of the C values for subject i, then this can be written in matrix notation as

$$\boldsymbol{\pi}_i^{(0)} = \mathbf{P}_i^{(0)} \boldsymbol{\pi}_i^{(0)}, \tag{5}$$

where $\boldsymbol{\pi}_i^{(0)} = (\pi_{i0}^{(0)}, \pi_{i1}^{(0)}, \ldots, \pi_{i,C-1}^{(0)})'$ and $\mathbf{P}_i^{(0)}$ is the $C \times C$ transition matrix with k–jth element $P(Y_{i0} = k | Y_{i,-1} = j)$. If we can find the value of $\boldsymbol{\pi}_i^{(0)}$ that solves Equation (5) for each subject, then we have found the 'optimal' values of these. Note that the transition probability values $P(Y_{i0} = k | Y_{i,-1} = j)$ can be obtained through the OLR model. One way to find the values of $\boldsymbol{\pi}_i^{(0)}$ is to find the dominant eigenvector of the transition matrix for each subject, $\mathbf{P}_i^{(0)}$ [21]. However, given that the E–M algorithm proceeds iteratively, and that we do not necessarily need to find an 'exact' solution to $\boldsymbol{\pi}_i^{(0)} = \mathbf{P}_i^{(0)} \boldsymbol{\pi}_i^{(0)}$ during the iteration procedure, it is possible to update the estimate of $\boldsymbol{\pi}_i^{(0)}$ at each iteration by replacing it with $\mathbf{P}_i^{(0)} \boldsymbol{\pi}_i^{(0)}$.

The `polr()` procedure will produce standard errors of the parameter estimates at the end of the E–M iteration procedure that are too small, as they have not been calculated under the mixture

model associated with $p(y_1)$. As in any GLM, the `polr()` procedure will obtain the standard errors based on the inverse of the information matrix. $\mathbf{I}(\boldsymbol{\Omega}) = -\partial^2 \log L(\boldsymbol{\Omega})/\partial\boldsymbol{\Omega}\partial\boldsymbol{\Omega}'$. While it is possible to evaluate analytically the required correction to the information matrix using the missing information principle [12], it is computationally easier to evaluate the information matrix numerically, i.e. by calculating

$$\frac{\partial^2 \log L(\boldsymbol{\Omega})}{\partial \Omega_i^2} \approx \frac{[\log L(\hat{\boldsymbol{\Omega}} + \boldsymbol{\delta}_i) - 2\log L(\hat{\boldsymbol{\Omega}}) + \log L(\hat{\boldsymbol{\Omega}} - \boldsymbol{\delta}_i)]}{h_i^2}$$

and

$$\frac{\partial^2 \log L(\boldsymbol{\Omega})}{\partial \Omega_i \Omega_j} \approx \frac{[\log L(\hat{\boldsymbol{\Omega}} + \boldsymbol{\delta}_i + \boldsymbol{\delta}_j) - \log L(\hat{\boldsymbol{\Omega}} + \boldsymbol{\delta}_i - \boldsymbol{\delta}_j) - \log L(\hat{\boldsymbol{\Omega}} - \boldsymbol{\delta}_i + \boldsymbol{\delta}_j) + \log L(\hat{\boldsymbol{\Omega}} - \boldsymbol{\delta}_i - \boldsymbol{\delta}_j)]}{4h_i h_j},$$

where $L(\hat{\boldsymbol{\Omega}})$ is the likelihood, evaluated at the maximum-likelihood estimates, $\hat{\boldsymbol{\Omega}}$, with $\boldsymbol{\delta}_i = (0, 0, \ldots, 0, h_i, 0, \ldots, 0)'$ and the differentials h_i chosen as $h_i = \hat{\Omega}_i \times 10^{-4}$.

The R code for all methods is available as a text file on the website http://www.kathryn.bartimote.com/research.

3. Example

The example data are a sub-set of results from a long-term cocoa (*Theobroma cocao* L.) trial conducted in Ladongi in the southern part of the island of Sulawesi, Indonesia. In many areas of south-east Asia, severe economic loss of cocoa and other crops occurs due to infection by the fungal pathogen *Phytophthora palmivora* [5]. The most common symptoms of this disease are cankers on the vertical stem and branches, and pod rot. A combination of chemical treatment as well as improved management practices (such as maintaining canopy ventilation, good drainage, sanitation, and frequent harvesting) can potentially help to decrease disease incidence due to *Phytophthora* infection and thereby have a positive effect on yield. In the experiment outlined here, researchers were assessing the effect of injections of potassium phosphonate (a derivative of phosphonic acid) into cocoa tree stems, coupled with improved management practices.

Treatment 1 (the control) consisted of water injection with improved management and Treatment 2 consisted of phosphonate injection with improved management. Each of the treatment groups consists of 50 trees. The cocoa trees are planted under shade trees in rows of approximately 20–25 trees that are spaced approximately 4 m apart. *Phytophthora* inoculum was not added at the beginning of the trial as trees were already naturally infected. The two treatments were applied alternately to trees. The systematic application of treatments is a limitation of the research since almost all statistical techniques assume random application of treatments to experimental units.

Infection was measured in a variety of ways. In this paper, we only refer to the canker severity ratings. A score of 0 indicates no canker, a score of 5 indicates the tree has been killed by canker, and scores 1–4 relate to active canker of increasing diameters by visual assessment. Trees were rated towards the end of each month by the same person, and data from the first 36 time points of the trial are used here. However, as there were only 12 trees that died (rating of 5), these were collapsed into the rating of 4, hence $C = 5$ ratings were used in the subsequent analysis.

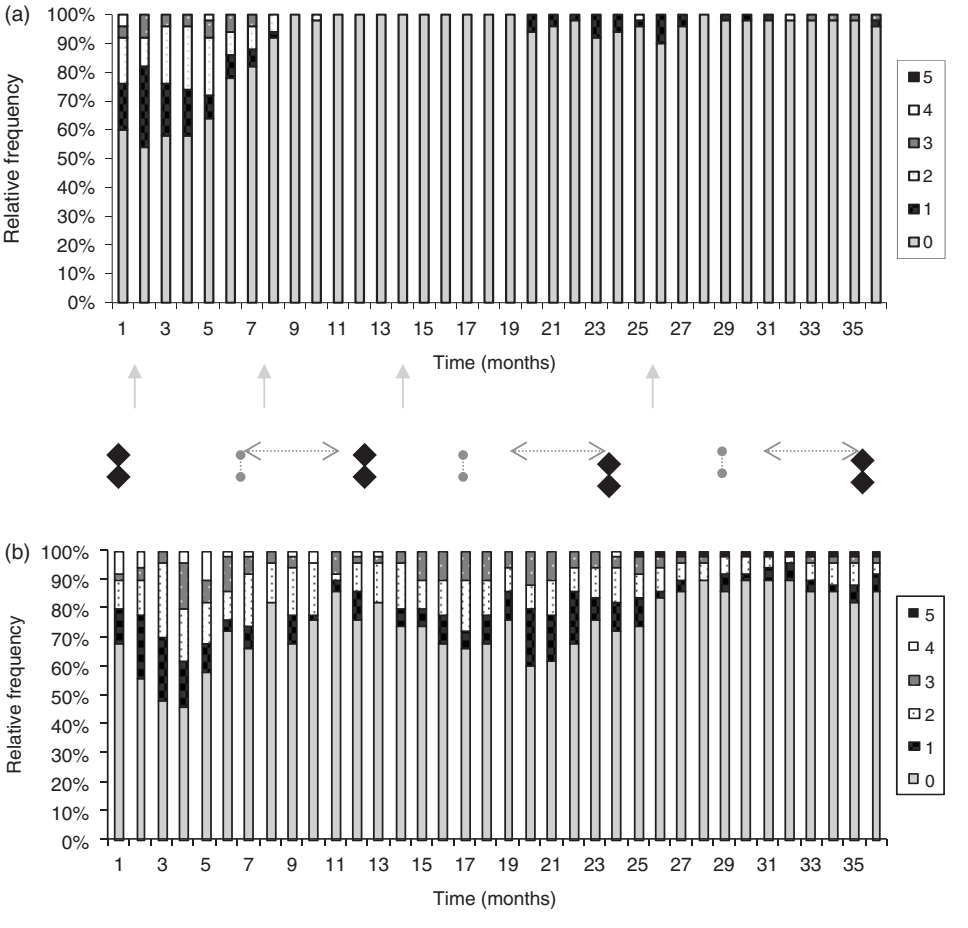

Figure 1. (a) Relative frequency of ratings for the phosphonate injection treatment. (b) Relative frequency of ratings for the water injection treatment.

3.1 *Data exploration*

The relatively large number of time points ($T = 36$) per tree (subject) provides us with sufficient information to be able to distinguish changes over time within individual trees from differences among trees in their baseline levels [4].

Figure 1 provides a summary of the data showing the relative frequency of each rating at each time point. The timing of the treatment injections and pod harvests, as well as the occurrence and length of the wet season, is shown. The overall distribution of ratings for the control and treatment groups is shown in Table 1.

In this experiment, severity ratings have been collected over a large number of relatively evenly spaced time points (measurements were taken during either mid-month or month-end). Measurement days were 30.7 days apart on average (SD = 3.8 days), with the range being from 22 to 40 days.

The initial (and full) model fitted to the data (for all approaches outlined in Section 2.1) is

$$\text{logit } P(Y_{it} \le k | Y_{i,t-1} = y_{i,t-1}) = \theta_k - \Big[\beta_1 \text{Time} + \beta_2 \text{Phosphonate} + \beta_3 \text{Time.Phosphonate} + \beta_4 \sin\left(\frac{2\pi \text{ Time}}{12}\right) + \beta_5 \cos\left(\frac{2\pi \text{ Time}}{12}\right) + \alpha' y^*_{i,t-1}\Big], \quad (6)$$

Table 1. Tally of ratings for each treatment aggregated over all times.

Rating	Water		Phosphonate		Total
0	1339	(74%)	1641	(91%)	2980
1	146	(8%)	76	(4%)	222
2	190	(11%)	58	(3%)	248
3	91	(5%)	22	(1%)	113
4	22	(1%)	3	(<1%)	25
5	12	(1%)	0	(0%)	12
N	1800	(100%)	1800	(100%)	3600

Table 2. Summary of fitted models for the cocoa canker rating data.

	Method 0				Method 5		
Model	Model df	AIC	X^2	P	AIC	X^2	P
Full	13	2627.90			2865.29		
– Phosphonate.Time	12	2630.09	4.19	0.041	2891.08	27.79	<0.001
– (SinTime + CosTime)	11	2632.20	8.30	0.016	2868.83	7.53	0.023
– y.star	9	4136.55	1516.65	<0.001	4366.76	1509.47	<0.001

Notes: The full model is shown with subsequent models having the particular term removed from the full model. 'Method 0' involves conditioning on data at time $t = 1$. 'Method 5' utilises the data at time $t = 1$ and is a model-based method for updating of $\pi_{ik}^{(0)}$. Because these two methods use different sets of data, the AIC are not comparable across methods.

where *Time* is the time elapsed (months: 1, 2, …, 36) since recording started, and *Phosphonate* is an indicator variable to distinguish treatment (1) from control (0). The sine and cosine terms have been included as an approximate way of capturing the influences of wet season timing and cocoa pod harvesting which are evident in the empirical ratings over time (Figure 1). The statistical significance of terms added to the model was checked by chi-square deviance difference tests (likelihood ratio tests) with a level of significance of 5%.

3.2 *Conditioning on the first observation*

The approach for the analysis here was to ignore $p(y_{i1})$ and effectively condition on y_{i1}. That is, a simple lag model was employed where response data lagged by one time point becomes a predictor variable in the model, using data at times $t = 2, \ldots, T$ only.

Deviance difference tests (likelihood ratio tests) were performed to see which (if any) terms could be dropped from the full model as defined in Equation (6). In addition, the Akaike information criterion (AIC) was determined for each fitted model, and these results are shown in Table 2, along with those for the model-based method (described below). As can be seen from the table, all terms in the model are significant, and hence we will retain the full model. The dependence on the previous observation is particularly strong ($P < 0.001$). From the estimated regression coefficients for $\mathbf{y}_{t-1}^{*}$, we see that the log odds for obtaining a rating of k or less at time t, given it had a rating of 0, 1, 2, or 3 at time $t - 1$ are 9.52, 6.81, 5.25, and 2.50 more respectively, compared with a tree with a rating of 4 at time $t - 1$ (based on $\sum_{j=k}^{C-2} \hat{\alpha}_j$).

The parameter estimates and standard errors for this method and the model-based method are shown in Table 3. For Method 0 (i.e. conditioning on the first observation), we see that ratings for trees injected with the phosphonate treatment decrease significantly more quickly over time than do the ratings for trees injected with the water (control) treatment as indicated by the significant *Time. Phosphonate* interaction term ($P = 0.041$), supporting the overall trend

Table 3. Parameter estimates and their standard errors of fitted models for the cocoa canker rating data.

	All 36 time points					First 6 time points only				
	Method 0		Method 5			Method 0		Method 5		
Model term	Coeff.	S.E.	Coeff.	S.E. (unadj.)	S.E (adj.)	Coeff.	S.E.	Coeff.	S.E. (unadj.)	S.E (adj.)
Phosphonate	−0.57	0.223	−0.41	0.205	0.218	−0.56	0.62	−0.11	0.44	0.53
Time	−0.02	0.007	−0.02	0.007	0.007	−2.33	1.23	−1.11	0.58	0.66
Phosphonate.Time	−0.03	0.014	−0.04	0.013	0.014	0.05	0.16	−0.04	0.12	0.14
SinTime	0.23	0.086	0.23	0.086	0.086	−0.45	1.38	1.01	0.46	0.59
CosTime	0.08	0.087	0.10	0.082	0.086	−4.83	2.42	−2.68	1.35	1.46
y.star1	−2.71	0.159	−2.69	0.154	0.158	−1.80	0.27	−1.77	0.25	0.27
y.star2	−1.56	0.183	−1.55	0.178	0.182	−1.95	0.32	−1.96	0.29	0.31
y.star3	−2.76	0.262	−2.78	0.256	0.264	−2.61	0.47	−2.62	0.43	0.48
y.star4	−2.50	0.449	−2.51	0.425	0.448	−1.53	0.62	−1.64	0.55	0.64
Intercept 0|1	−7.31	0.495	−7.29	0.469	0.489	−14.03	4.98	−9.06	2.18	2.53
Intercept 1|2	−6.13	0.490	−6.09	0.463	0.483	−12.67	4.97	−7.69	2.17	2.52
Intercept 2|3	−3.57	0.461	−3.53	0.434	0.453	−10.15	4.96	−5.15	2.15	2.49
Intercept 3|4	−0.56	0.367	−0.65	0.348	0.355	−7.63	4.92	−2.95	2.11	2.43

Notes: 'Method 0' involves conditioning on data at time $t = 1$. 'Method 5' utilises the data at time $t = 1$ and is a model-based method for updating of $\pi_{ik}^{(0)}$. The standard errors shown for 'Method 0' and the unadjusted ('unadj') ones for 'Method 5' are from the standard `polr()` output, whereas the ones labelled 'adj.' are the adjusted ones by re-computing the inverse of the information matrix.

shown in Figure 1. In particular, the log odds of a certain score or lower for a control plant increases at a rate of 0.02 per month (se = 0.007) while that for a treated plant increases at a significantly faster rate of 0.05 per month (se = 0.012). (The slope of the phosphonate treatment group is estimated as $\hat{\beta}_T + \hat{\beta}_{T.P}$ (T: time, P: phosphonate) with an estimated standard error $\left[\text{se}^2(\hat{\beta}_T) + \text{se}^2(\hat{\beta}_{T.P}) + 2r(\hat{\beta}_T, \hat{\beta}_{T.P})\text{se}(\hat{\beta}_T)\text{se}(\hat{\beta}_{T.P})\right]^{1/2}$, with the standard errors $\text{se}(\hat{\beta}_T)$ and $\text{se}(\hat{\beta}_{T.P})$ shown on Table 3, and correlation $r(\hat{\beta}_T, \hat{\beta}_{T.P}) = -0.472$). From the findings above, we also know that the rating for a particular tree depends heavily on the rating given to that tree in the previous time period ($P < 0.001$), and that there is an overall seasonal pattern to the severity of disease symptoms ($P = 0.016$).

3.3 *Fitting the full model using estimates of* $\pi_{ik}^{(0)}$

Sections 2.3–2.5 outline the methods employed here. In the cocoa disease severity example, the use of a mixture distribution to handle the unknown outcome at $t = 0$ to assist in calculating the full log likelihood is appropriate since disease was already present in the trees before they were included in the experiment and measurement began at what is notionally labelled 'time 1'. In the interests of brevity, we will only provide the results of the model-based method here (Method 5) which involves the iterative updating of estimates of $\pi_{ik}^{(0)}$.

We see in Table 2 that again all terms in the model are significant. It should also be noted that the significance of the Phosphonate × Time interaction is now much stronger, than detected using 'Method 0'. Table 3 (left-hand side) also includes the estimated regression coefficients and their standard errors. In general, the regression coefficients are in agreement with those using the simple transition model (Method 0: conditioning on the first observation).

Note that the standard errors are almost always smaller than those for Method 0 since we have an additional set of observations included in the model (from time 1). However they are a little too small, being the ones provided from a standard `polr()` call, and should strictly be adjusted via

Table 4. Comparison of observed and expected frequencies of ratings.

	Water		Phosphonate		Total	
Rating	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	1339	1325.9	1641	1645.7	2980	2971.6
1	146	151.0	76	66.9	222	217.9
2	190	193.1	58	63.3	248	256.4
3	91	95.8	22	19.9	113	115.7
4	34	34.2	3	4.2	37	38.3
Total	1800	1800.0	1800	1800.0	3600	3600.0

Note: Expected frequencies have been obtained from summing all the fitted event probabilities for each rating value, across all 1800 observations within the water- and phosphonate-treated trees.

methods such as those described in Section 2.5. These adjusted standard errors are shown in the last column of Table 3. For this example data set, there is no great benefit of computing the adjusted standard errors. However, in other situations, where the time series is considerably shorter, there may be a larger effect of adjustment, due to the relatively large amount of 'missing information'.

Some indication of the goodness-of-fit can be obtained from the residual deviance statistic of 2839 on 3587 degrees of freedom. (The residual deviance is calculated as $-2\log L(\hat{\boldsymbol{\Omega}})$, since the log-likelihood for the 'saturated model' is zero, for unit record data.) This would indicate some under-dispersion, although this cannot be compared with a chi-square distribution, being based on unit record data. However, a further indication of fit can be obtained by comparing the observed frequency distribution of ratings to the expected values within the control and treatment groups, based on the fitted model. The expected frequencies have been obtained as $E_{kG} = \sum_{i\in G} P(Y_{i1} = k) + \sum_{i\in G}\sum_{t=2}^{T} P(Y_{it} = k|Y_{i,t-1} = y_{i,t-1}),\ k = 0, 1, \ldots, C-1,\ G =$ treatment group, evaluated at $\boldsymbol{\Omega} = \hat{\boldsymbol{\Omega}}$. The comparison shown in Table 4 would indicate a high level of agreement: a Pearson chi-square statistic for this is 2.83 on 8 df, although again cannot be compared to a chi-square distribution.

3.4 *Re-analysis using a short time series*

To investigate the potential benefits of the model-based method for shorter time series, the first six of the 36 time points available were taken as a second data set. These data were analysed via both the simple model (conditioning on time 1, Method 0), and the model-based method (Method 5). The results are given in Table 3 on the right-hand side.

As can be seen there is a marked reduction (up to 50% or more) in almost all the standard errors of parameter estimates using Method 5, compared with Method 0, which demonstrates the substantial improvement for a short time series when information from time $t = 1$ is fully utilised, as opposed to conditioning on these data values. It also demonstrates the need to adjust the standard errors for short time series, with unadjusted ones typically being $\sim$13% too small.

4. Discussion

The example data set has demonstrated how the methods developed here (based on standard OLR methods) can be adapted to address a commonly encountered problem in agricultural and horticultural research, although the area of application is far broader. The focus of the method has primarily been on how to draw inference about the experiment, so in the current example, it was found that ratings for trees injected with the phosphonate treatment decrease significantly more quickly over time than do the ratings for trees injected with the water (control) treatment.

Also the rating for a particular tree depends heavily on the rating given to that tree in the previous time period, and that there is an overall seasonal pattern to the severity of disease symptoms with more severe disease apparent during the wet season of each year (Figure 1). We acknowledge that it may be interesting to explore higher-order Markov models (i.e. dependence on observations further removed than the adjacent time period, e.g. $t - 2$) for these data. Further, interactions between the previous observation and other predictor variables could be included in the model; or in the case of higher-order Markov models, interactions between responses for previous time points and/or interactions between previous responses and other predictor variables. However, the choice to proceed with the first order without the additional model complexities allowed us to simplify explanations of the methodology as it developed.

Here the effect of the previous time incorporates both the serial correlation (time) and the cluster correlation (tree effect). In this example, the purpose of the research was not to model or fully understand the correlation structure, but to test for a treatment effect (and ensure that any correlation is accounted for in order to gain an accurate view of the treatment effect). However, if we wanted to estimate the disease status (i.e. cumulative probability distribution) of an individual tree, a model incorporating a random term for the tree would be appropriate. Alternatively, if we wanted to model the cluster effect of the trees from the two treatment groups separately to the serial correlation, marginal models (or GEEs) would be suitable.

The current data set consisted of relatively long time series (36 observations per tree). For this situation, there is only a marginal improvement when changing from a simple conditional model to a model-based method, as demonstrated by only a slight reduction in standard errors of parameter estimates. However, in the situation considered here with a relatively short time series based on the first six observations only, the benefit was demonstrated to be much more substantial, warranting the additional complexity of fitting such a model. Clearly, additional research needs to be conducted to further evaluate the merits of the different methods, in particular for assessing more fully the effect of shorter time series, by making use of simulation studies.

Additionally, the example data set has a large number of evenly spaced time points. More complex models would be required to cater for irregularly spaced observation times. For example, the transition model would not be directly applicable for purely randomly observed times as the model assumes the dependence is on the previous time(s) spaced at unit time intervals.

Recently, many agronomic and horticultural field studies have addressed spatial correlation in their analyses, at least for the case of normally distributed response data, and some important methodology has been developed in this area [23]. This would be of relevance in the current application, given the grid layout of the cocoa trees. In theory, it would be possible to extend the Markov model to include information on disease ratings of adjacent trees, perhaps at previous points of time.

In summary, a method is presented for the analysis of ordinal time series data that in its simplest form (conditioning on the first observation) can be readily implemented in any software that provides OLR. Extensions are provided to make full use of information from the first observation by treating the rating at time $t = 0$ as a missing data problem. However, by use of the E–M algorithm, even this more complex model is readily incorporated into a standard ordinal logistic model framework, provided the package has the capability of iteration and of specifying case weights, as is offered in the R computing environment.

Acknowledgements

We would like to thank Professor David Guest from The University of Sydney and Dr Peter J. McMahon from La Trobe University and their colleagues in Indonesia, Abdul Wahab and Andi Aryadi for the provision of the example data. KBA would also acknowledge the Faculty of Agriculture, Food and Natural Resources for their financial support of this work, while she was a postgraduate research student there.

References

[1] A. Agresti, *Categorical Data Analysis*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2002.
[2] J.E. Corrente and M.D.P. Diaz, *Ordinal models and generalized estimating equations to evaluate disease severity*, J. Appl. Stat. 30 (2003), pp. 425–439.
[3] A.P. Dempster, N.M. Laird, and D.A. Rubin, *Maximum likelihood estimation from incomplete data via the EM algorithm* (*with discussion*), J. R. Stat. Soc. Ser. B 39 (1977), pp. 1–38.
[4] P.J. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*, 2nd ed., Oxford University Press, Oxford, 2002.
[5] A. Drenth and D.I. Guest, *Diversity and Management of Phytophthora in Southeast Asia*, Australian Centre for International Agricultural Research, Canberra, 2004.
[6] G.M. Fitzmaurice, N.M. Laird, and A.G. Rotnitzky, *Regression models for discrete longitudinal responses*, Statist. Sci. 8 (1993), pp. 284–299.
[7] V.E. Johnson and J.H. Albert, *Ordinal Data Modeling*, Springer, New York, 1999.
[8] C. Lange and J.C. Whittaker, *Mapping quantitative trait loci using generalized estimating equations*, Genetics 159 (2001), pp. 1325–1337.
[9] K. Lee and M.J. Daniels, *A class of Markov models for longitudinal ordinal data*, Biometrics 63 (2007), pp. 1060–1067.
[10] K.-Y. Liang and S.L. Zeger, *Longitudinal data analysis using generalized linear models*, Biometrika 78 (1986), pp. 13–22.
[11] J. Lindsey, B. Jones, and A.F. Ebbutt, *Simple models for repeated ordinal responses with an application to a seasonal; rhinitis clinical trial*, Stat. Med. 16 (1997), pp. 2873–2882.
[12] T.A. Louis, *Finding the observed information matrix when using the EM algorithm*, J. R. Stat. Soc. Ser. B Methodol. 44 (1982), pp. 226–233.
[13] P. McCullagh, and J.A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman & Hall, London, 1989.
[14] C.E. McCulloch and S.R. Searle, *Generalized, Linear, and Mixed Models*, John Wiley & Sons, New York, 2001.
[15] G.J. McLachlan and K.E. Basford, *Mixture Models: Inference and Applications to Clustering*, *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, 1988.
[16] G. Molenberghs, M.G. Kenward, and E. Lesaffre, *The analysis of longitudinal ordinal data with nonrandom drop-out*, Biometrika 84 (1997), pp. 33–44.
[17] G. Molenberghs and E. Lesaffre, *Marginal modeling of correlated ordinal data using a multivariate Plackett distribution*, J. Amer. Statist. Assoc. 89 (1994), pp. 633–644.
[18] N.R. Parsons, R.N. Edmondson, and S.G. Gilmour, *A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research*, Appl. Stat. 55 (2006), pp. 507–524.
[19] H.D. Patterson and R. Thompson, *Recovery of inter-block information when block sizes are unequal*, Biometrika 58 (1971), pp. 545–554.
[20] R Development Core Team, *R*: *A Language and Environment for Statistical Computing* (Version 2.8.1) [for Windows], Vienna, Austria, 2008. Available at http://www.r-project.org/.
[21] S.R. Searle, *Matrix Algebra Useful for Statistics*, Wiley, New York, 1982.
[22] D.A. Shah and L.V. Madden, *Nonparametric analysis of ordinal data in designed experiments*, Phytopathology 94 (2004), pp. 33–43.
[23] A. Smith, B. Cullis, and R. Thompson, *Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend*, Biometrics 57 (2001), pp. 1138–1147.
[24] P.C. Thomson, *A generalized estimating equations approach to quantitative trait locus detection of non-normal traits*, Genet. Sel. Evol. 35 (2003), pp. 257–280.
[25] F. Yu, H. Morgenstern, E. Hurwitz, and T.R. Berlin, *Use of a Markov transition model to analyse longitudinal low-back pain data*, Stat. Methods Med. Res. 12 (2003), pp. 321–331.
[26] S.L. Zeger and K.-Y. Liang, *An overview of methods for the analysis of longitudinal data*, Stat. Med. 11 (1992), pp. 1825–1839.

Appendix

A.1 Derivation of the estimating equations for estimating parameters

The expression for the log-likelihood (3) is

$$\log L(\boldsymbol{\Omega}) = \sum_{i=1}^{n} \log \sum_{k=0}^{C-1} \pi_{ik}^{(0)} p(y_{i1}|y_{i0} = k) + \sum_{i=1}^{n} \sum_{t=2}^{T} \log p(y_{it}|y_{i,t-1}).$$

Hence, the estimating equations to be solved are $\partial \log L(\boldsymbol{\Omega})/\partial \boldsymbol{\Omega} = \mathbf{0}$, where

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega}} &= \sum_{i=1}^{n} \frac{\sum_{k=0}^{C-1} \left[(\partial \pi_{ik}^{(0)}/\partial \boldsymbol{\Omega})\, p(y_{i1}|y_{i0}=k) + \pi_{ik}^{(0)} (\partial/\partial \boldsymbol{\Omega})\, p(y_{i1}|y_{i0}=k) \right]}{\sum_{k=0}^{C-1} \pi_{ik}^{(0)} p(y_{i1}|y_{i0}=k)} \\
&\quad + \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{it}|y_{i,t-1}) \\
&= \sum_{i=1}^{n} \frac{\sum_{k=0}^{C-1} \pi_{ik}^{(0)} p(y_{i1}|y_{i0}=k) \left[(\partial/\partial \boldsymbol{\Omega}) \log \pi_{ik}^{(0)} + (\partial/\partial \boldsymbol{\Omega}) \log p(y_{i1}|y_{i0}=k) \right]}{\sum_{k=0}^{C-1} \pi_{ik}^{(0)} p(y_{i1}|y_{i0}=k)} \\
&\quad + \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{it}|y_{i,t-1}) \\
&= \sum_{i=1}^{n} \sum_{k=0}^{C-1} \tau_{ik}^{(0)} \frac{\partial}{\partial \boldsymbol{\Omega}} \log \pi_{ik}^{(0)} + \sum_{i=1}^{n} \sum_{k=0}^{C-1} \tau_{ik}^{(0)} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{i1}|y_{i0}=k) \\
&\quad + \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{\partial}{\partial \boldsymbol{\Omega}} \log p(y_{it}|y_{i,t-1})
\end{aligned}$$

and where

$$\tau_{ik}^{(0)} = \frac{\pi_{ik}^{(0)} p(y_{i1}|y_{i0}=k)}{\sum_{k'=0}^{C-1} \pi_{ik'}^{(0)} p(y_{i1}|y_{i0}=k')} = \frac{\pi_{ik}^{(0)} p(y_{i1}|y_{i0}=k)}{p(y_{i1})} = p(y_{i0}=k|y_{i1}).$$

So the estimating function $\partial \log L(\boldsymbol{\Omega})/\partial \boldsymbol{\Omega}$ consists of: a weighted term for the 'mixing proportions', $\pi_{ik}^{(0)}$, a weighted term for the conditional probabilities, $p(y_{i1}|y_{i0}=k)$, and unweighted term for all the subsequent conditional probabilities, $p(y_{it}|y_{i,t-1})$ (which may be considered to have weights all equal to one). The E–M algorithm progresses at each iteration by using a weighted OLR analysis corresponding to the second two terms, while the first term is handled separately (by a separate procedure described in Section 2.5) to calculate the $\pi_{ik}^{(0)}$).

A.2 Computation of the weights, $\tau_{ik}^{(0)}$, based on empirical proportions

Method 1: Based on overall empirical proportions, using the overall empirical proportions for each rating level, regardless of tree, time or treatment, i.e.

$$\hat{\pi}_{ik}^{(0)} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} I(Y_{it}=k),$$

where $I(\cdot)$ is the 0–1 indicator function.

Method 2: Based on overall empirical proportions at $t=1$, using the empirical proportions for each rating level, based on time $t=1$ only, regardless of tree or treatment, i.e.

$$\hat{\pi}_{ik}^{(0)} = \frac{1}{n} \sum_{i=1}^{n} I(Y_{i1}=k).$$

Method 3: Based on tree-specific empirical proportions, using the empirical proportions for each rating level, based on tree = i, regardless of time, i.e.

$$\hat{\pi}_{ik}^{(0)} = \frac{1}{T}\sum_{t=1}^{T} I(Y_{it} = k).$$

Method 4: Assume that the rating at time 0 is the same as that at time 1, for each tree, i.e.

$$\hat{\pi}_{ik}^{(0)} = I(Y_{i1} = k).$$