

# Regression Models for Ordinal Categorical Time Series Data

Brajendra C. Sutradhar and R. Prabhakar Rao

**Abstract** Regression analysis for multinomial/categorical time series is not adequately discussed in the literature. Furthermore, when categories of a multinomial response at a given time are ordinal, the regression analysis for such ordinal categorical time series becomes more complex. In this paper, we first develop a lag 1 transitional logit probabilities based correlation model for the multinomial responses recorded over time. This model is referred to as a multinomial dynamic logits (MDL) model. To accommodate the ordinal nature of the responses we then compute the binary distributions for the cumulative transitional responses with cumulative logits as the binary probabilities. These binary distributions are next used to construct a pseudo likelihood function for inferences for the repeated ordinal multinomial data. More specifically, for the purpose of model fitting, the likelihood estimation is developed for the regression and dynamic dependence parameters involved in the MDL model.

**Keywords** Category transition over time · Cumulative logits · Marginal multinomial logits · Multinomial dynamic logits · Pseudo binary likelihood

## 1 Introduction

There are situations in practice where a univariate multinomial response, for example, the economic profit status of a pharmaceutical industry such as poor, medium, or high, may be recorded over the years along with known covariates such as type of industry, yearly advertising cost, and other research and development expenditures. It is likely that the profit status of an industry in a given year is correlated with status of profits from the past years. It is of interest to know both (i) the effects of the time dependent

---

B.C. Sutradhar (✉)  
Memorial University, St. John's A1C5S7, Canada  
e-mail: bsutradh@mun.ca

R. Prabhakar Rao  
Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, Anantapur, India  
e-mail: rprabhakarao@gmail.com

covariates, and (ii) the dynamic relationship among the responses over the years. This type of multinomial time series data has been analyzed by some authors such as Fahrmeir and Kaufmann [4], Kaufmann [8], Fokianos and Kedem [5–7], and Loredo-Osti and Sutradhar [10]. As far as the dynamic relationship is concerned, Loredo-Osti and Sutradhar [10] have considered a multinomial dynamic logit (MDL) model as a generalization of the binary time series model used in Tagore and Sutradhar [16] (see also Tong [17]).

Suppose that  $y_t = (y_{t1}, \dots, y_{tj}, \dots, y_{t,J-1})'$  denotes the  $(J - 1)$ -dimensional multinomial response variable and for  $j = 1, \dots, J - 1$ ,

$$y_t^{(j)} = (y_{t1}^{(j)}, \dots, y_{tj}^{(j)}, \dots, y_{t,J-1}^{(j)})' = (01'_{j-1}, 1, 01'_{j-1-j})' \equiv \delta_{tj} \tag{1}$$

indicates that the multinomial response recorded at time  $t$  belongs to the  $j$ th category. For  $j = J$ , one writes  $y_t^{(J)} = \delta_{tJ} = 01_{J-1}$ . Here and also in (1), for a scalar constant  $c$ , we have used  $c1_j$  for simplicity, to represent  $c \otimes 1_j$ ,  $\otimes$  being the well known Kronecker or direct product. This notation will also be used through out the rest of the paper when needed. Note that in the non-stationary case, that is, when covariates are time dependent, one uses the time dependent marginal probabilities. Specifically, suppose that at time point  $t$  ( $t = 1, \dots, T$ ),  $x_t = (x_{t1}, \dots, x_{t\ell}, \dots, x_{t,p+1})'$  denotes the  $(p + 1)$ -dimensional covariate vector and  $\beta_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})'$  denotes the effect of  $x_t$  on  $y_t^{(j)}$  for  $j = 1, \dots, J - 1$ , and all  $t = 1, \dots, T$ ,  $T$  being the length of the time series. In such cases, the multinomial probability at time  $t$ , has the form

$$P[y_t = y_t^{(j)}] = \pi_{(t)j} = \begin{cases} \frac{\exp(x_t' \beta_j)}{1 + \sum_{g=1}^{J-1} \exp(x_t' \beta_g)} & \text{for } j = 1, \dots, J - 1; \quad t = 1, \dots, T \\ \frac{1}{1 + \sum_{g=1}^{J-1} \exp(x_t' \beta_g)} & \text{for } j = J; \quad t = 1, \dots, T, \end{cases} \tag{2}$$

and the elements of  $y_t = (y_{t1}, \dots, y_{tj}, \dots, y_{t,J-1})'$  at time  $t$  follow the multinomial probability distribution given by

$$P[y_{t1}, \dots, y_{tj}, \dots, y_{t,J-1}] = \prod_{j=1}^J \pi_{(t)j}^{y_{tj}}, \tag{3}$$

for all  $t = 1, \dots, T$ . In (3),  $y_{tJ} = 1 - \sum_{j=1}^{J-1} y_{tj}$ , and  $\pi_{tJ} = 1 - \sum_{j=1}^{J-1} \pi_{tj}$ .

Next we define the transitional probability from the  $g$ th ( $g = 1, \dots, J$ ) category at time  $t - 1$  to the  $j$ th category at time  $t$ , given by

$$\begin{aligned} \eta_{t|t-1}^{(j)}(g) &= P\left(Y_t = y_t^{(j)} \mid Y_{t-1} = y_{t-1}^{(g)}\right) \\ &= \begin{cases} \frac{\exp\left[x_t' \beta_j + \gamma_j' y_{t-1}^{(g)}\right]}{1 + \sum_{v=1}^{J-1} \exp\left[x_t' \beta_v + \gamma_v' y_{t-1}^{(g)}\right]}, & \text{for } j = 1, \dots, J - 1 \\ \frac{1}{1 + \sum_{v=1}^{J-1} \exp\left[x_t' \beta_v + \gamma_v' y_{t-1}^{(g)}\right]}, & \text{for } j = J, \end{cases} \end{aligned} \tag{4}$$

where  $\gamma_j = (\gamma_{j1}, \dots, \gamma_{jv}, \dots, \gamma_{j,J-1})'$  denotes the dynamic dependence parameters. Note that this model in (4) is referred to as the multinomial dynamic logits (MDL) model. For the binary case ( $J = 2$ ), this type of non-linear dynamic logit model has been studied by some econometricians. See, for example, Amemiya [3, p. 422] in time series setup, and the recent book by Sutradhar [13, Sect. 7.7] in the longitudinal setup. Now for further notational convenience, we re-express the conditional probabilities in (4) as

$$\eta_{t|t-1}^{(j)}(g) = \begin{cases} \frac{\exp[x'_t \beta_j + \gamma'_j \delta_{(t-1)g}]}{1 + \sum_{v=1}^{J-1} \exp[x'_t \beta_v + \gamma'_v \delta_{(t-1)g}]}, & \text{for } j = 1, \dots, J-1 \\ \frac{1}{1 + \sum_{v=1}^{J-1} \exp[x'_t \beta_v + \gamma'_v \delta_{(t-1)g}]}, & \text{for } j = J, \end{cases} \quad (5)$$

where for  $t = 2, \dots, T$ ,  $\delta_{(t-1)g}$ , by (1), has the formula

$$\delta_{(t-1)g} = \begin{cases} [01'_{g-1}, 1, 01'_{J-1-g}]' & \text{for } g = 1, \dots, J-1 \\ 01_{J-1} & \text{for } g = J. \end{cases}$$

Remark that in (5), the category  $g$  occurred at time  $t - 1$ . Thus the category  $g$  depends on time  $t - 1$ , and  $\delta_{(t-1)g} \equiv \delta_{g_{t-1}}$ . However for simplicity we have used  $g$  for  $g_{t-1}$ .

Let  $\beta = (\beta'_1, \dots, \beta'_j, \dots, \beta'_{J-1})' : (p + 1)(J - 1) \times 1$ , and  $\gamma = (\gamma'_1, \dots, \gamma'_j, \dots, \gamma'_{J-1})' : (J - 1)^2 \times 1$ . These parameters are involved in the unconditional mean, variance and covariances of the responses. More specifically one may show [10] that

$$\begin{aligned} E[Y_t] &= \tilde{\pi}_{(t)}(\beta, \gamma) = (\tilde{\pi}_{(t)1}, \dots, \tilde{\pi}_{(t)j}, \dots, \tilde{\pi}_{(t)(J-1)})' : (J - 1) \times 1 \\ &= \begin{cases} [\pi_{(1)1}, \dots, \pi_{(1)j}, \dots, \pi_{(1)(J-1)}]' & \text{for } t = 1 \\ \eta_{(t|t-1)}(J) + [\eta_{(t|t-1),M} - \eta_{(t|t-1)}(J)1'_{J-1}] \tilde{\pi}_{(t-1)} & \text{for } t = 2, \dots, T-1 \end{cases} \end{aligned} \quad (6)$$

$$\begin{aligned} \text{var}[Y_t] &= \text{diag}[\tilde{\pi}_{(t)1}, \dots, \tilde{\pi}_{(t)j}, \dots, \tilde{\pi}_{(t)(J-1)}] - \tilde{\pi}_{(t)} \tilde{\pi}'_{(t)} \\ &= (\text{cov}(Y_{tj}, Y_{tk})) = (\tilde{\sigma}_{(tt)jk}), \quad j, k = 1, \dots, J-1 \\ &= \tilde{\Sigma}_{(tt)}(\beta, \gamma), \quad \text{for } t = 1, \dots, T \end{aligned} \quad (7)$$

$$\begin{aligned} \text{cov}[Y_u, Y_t] &= \Pi^t_{s=u+1} [\eta_{(s|s-1),M} - \eta_{(s|s-1)}(J)1'_{J-1}] \text{var}[Y_u], \quad \text{for } u < t, t = 2, \dots, T \\ &= (\text{cov}(Y_{uj}, Y_{tk})) = (\tilde{\sigma}_{(ut)jk}), \quad j, k = 1, \dots, J-1 \\ &= \tilde{\Sigma}_{(ut)}(\beta, \gamma), \end{aligned} \quad (8)$$

where

$$\begin{aligned} \eta_{(s|s-1)}(J) &= [\eta_{s|s-1}^{(1)}(J), \dots, \eta_{s|s-1}^{(j)}(J), \dots, \eta_{s|s-1}^{(J-1)}(J)]' = \pi_{(s)} : (J - 1) \times 1 \\ \eta_{(s|s-1),M} &= \begin{pmatrix} \eta_{s|s-1}^{(1)}(1) & \dots & \eta_{s|s-1}^{(1)}(g) & \dots & \eta_{s|s-1}^{(1)}(J-1) \\ \vdots & & \vdots & & \vdots \\ \eta_{s|s-1}^{(j)}(1) & \dots & \eta_{s|s-1}^{(j)}(g) & \dots & \eta_{s|s-1}^{(j)}(J-1) \\ \vdots & & \vdots & & \vdots \\ \eta_{s|s-1}^{(J-1)}(1) & \dots & \eta_{s|s-1}^{(J-1)}(g) & \dots & \eta_{s|s-1}^{(J-1)}(J-1) \end{pmatrix} : (J - 1) \times (J - 1). \end{aligned}$$

Notice that there is a relation between the vector  $\eta_{(s|s-1)}(J)$  and the matrix  $\eta_{(s|s-1),M}$ . This is because the transition matrix  $\eta_{(s|s-1),M}$  contains the transitional probabilities from any of the first  $J - 1$  states at time  $s - 1$  to any of the  $J - 1$  states at time  $s$ , whereas the transition vector  $\eta_{(s|s-1)}(J)$  contains transitional probabilities from the  $J$ th state at time  $s - 1$  to any of the first  $J - 1$  states at time  $s - 1$ . Consequently, once the transition matrix  $\eta_{(s|s-1),M}$  is computed, the transition vector  $\eta_{(s|s-1)}(J)$  becomes known.

It is of importance to estimate  $\beta$  and  $\gamma$  parameters mainly to understand the aforementioned basic properties including the pair-wise correlations of the responses.

Note however that the multinomial time series model (2)–(5) and its basic moment properties shown in (6)–(8) are derived without any order restrictions of the categories of the responses. The purpose of this paper is to estimate the parameters  $\beta$  and  $\gamma$  under an ordinal categorical response model which we describe in Sect. 2. In Sect. 3, we demonstrate the application of a pseudo likelihood approach for the estimation for these parameters. Some concluding remarks are made in Sect. 4.

## 2 Cumulative MDL Model for Ordinal Categorical Data

When categories for a response at a given time  $t$  are ordinal, one may then collapse the  $J > 2$  categories in a cumulative fashion into two ( $J' = 2$ ) categories and use simpler binary model to fit such collapsed data. Note however that there will be various binary groups depending on which category in the middle is used as a cut point. For the transitional categorical response from time  $t - 1$  (say) to time  $t$ , cumulation of the categories at time  $t$  has to be computed conditional on the cumulative categories at time  $t - 1$ . This will also generate a binary model for cumulative transitional responses. These concepts of cumulative probabilities for a cumulative response are used in the next section to construct the desired cumulative MDL model.

### 2.1 Marginal Cumulative Model at Time $t = 1$

Suppose that for a selected cut point  $j$  ( $j = 1, \dots, J - 1$ ),  $F_{(1)j} = \sum_{c=1}^j \pi_{(1)c}$  represents the probability for a multinomial response to be in category  $c$  between 1 and  $j$ , where  $\pi_{(1)c}$  by (2) defines the probability for the response to be in category  $c$  ( $c = 1, \dots, J$ ) at time  $t = 1$ . Thus,  $1 - F_{(1)j} = \sum_{c=j+1}^J \pi_{(1)c}$  would represent the probability for the multinomial response to be in category  $c$  beyond  $j$ . To reflect this binary nature of the observed response in category  $c$  with regard to cut point  $j$ , we define a binary variable  $b_c^{(j)}(1)$  such that

$$P[b_c^{(j)}(1) = 1] = 1 - F_{(1)j} = \sum_{c=j+1}^J \pi_{(1)c}. \quad (9)$$

Notice that because there are  $J - 1$  possible cut points, if the categories are ordered and the response falls in  $c$ th category, by (11) below, we then obtain the cut points based observed vector at time  $t = 1$  as

$$[b_c^{(1)}(1) = 1, \dots, b_c^{(c-1)}(1) = 1, b_c^{(c)}(1) = 0, \dots, b_c^{(J-1)}(1) = 0].$$

For other values of  $t$ , the observed responses are constructed similarly depending on the response category.

### 2.2 Lag 1 Transitional Cumulative Model at Time $t = 2, \dots, T$

In order to develop a transitional model, suppose we observe that the multinomial response at time  $t - 1$  ( $t = 2, \dots, T$ ) was in  $c_1$ th category ( $c_1 = 1, \dots, J$ ), whereas at time  $t$  it is observed in  $c_2$  ( $c_2 = 1, \dots, J$ ) category. Let  $(g, j)$  denote a bivariate cut point which facilitates the binary variables [similar to (9)] given by

$$b_{c_1}^{(g)}(t - 1) = \begin{cases} 1 & \text{for the response in category } c_1 > g \text{ at time } t - 1 \\ 0 & \text{for the response in category } c_1 \leq g \text{ at time } t - 1, \end{cases} \tag{10}$$

and

$$b_{c_2}^{(j)}(t) = \begin{cases} 1 & \text{for the response in category } c_2 > j \text{ at time } t \\ 0 & \text{for the response in category } c_2 \leq j \text{ at time } t. \end{cases} \tag{11}$$

Consequently, a transitional probability model based on conditional probabilities (5) may be written as

$$P[b_{c_2}^{(j)}(t) = 1 | b_{c_1}^{(g)}(t - 1)] = \tilde{\lambda}_{gj}^{(2)}(b_{c_1}^{(g)}(t - 1)) = \begin{cases} \tilde{\lambda}_{gj}^{(2)}(1) & \text{for } b_{c_1}^{(g)}(t - 1) = 0 \\ \tilde{\lambda}_{gj}^{(2)}(2) & \text{for } b_{c_1}^{(g)}(t - 1) = 1, \end{cases} \tag{12}$$

$$= \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1), \end{cases} \tag{13}$$

where the conditional probability  $\lambda_{t|t-1}^{(c_2)}(c_1)$ , has the known multinomial dynamic logit (MDL) form given by (5). For convenience, following (12)–(13), we also write

$$P[b_{c_2}^{(j)}(t) = 0 | b_{c_1}^{(g)}(t - 1)] = 1 - \tilde{\lambda}_{gj}^{(2)}(b_{c_1}^{(g)}(t - 1)) = \begin{cases} \tilde{\lambda}_{gj}^{(1)}(1) = 1 - \tilde{\lambda}_{gj}^{(2)}(1) & \text{for } b_{c_1}^{(g)}(t - 1) = 0 \\ \tilde{\lambda}_{gj}^{(1)}(2) = 1 - \tilde{\lambda}_{gj}^{(2)}(2) & \text{for } b_{c_1}^{(g)}(t - 1) = 1, \end{cases} \tag{14}$$

$$= \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \left[ 1 - \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) \right] \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \left[ 1 - \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) \right] \end{cases} \quad (15)$$

$$= \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=1}^j \lambda_{t|t-1}^{(c_2)}(c_1) \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=1}^j \lambda_{t|t-1}^{(c_2)}(c_1). \end{cases} \quad (16)$$

### 3 Pseudo Binary Likelihood Estimation for the Ordinal Model

In this section, we construct a binary data based likelihood function, where the binary data are obtained by collapsing the available ordinal multinomial observations. Consequently, we refer to this likelihood approach as the so-called pseudo likelihood approach. However, for convenience, we use the terminology ‘likelihood’ for the ‘pseudo likelihood’, through out the section.

At  $t = 1$ , the marginal likelihood for  $\beta$  by (9) has the form

$$\begin{aligned} L_1(\beta) &= \prod_{j=1}^{J-1} \left[ \{F_{(1)j}\}^{1-b_c^{(j)}(1)} \right] \left[ \{1 - F_{(1)j}\}^{b_c^{(j)}(1)} \right] \\ &= \prod_{j=1}^{J-1} \left[ \left\{ \sum_{c=1}^j \pi_{(1)c} \right\}^{1-b_c^{(j)}(1)} \right] \left[ \left\{ \sum_{c=j+1}^J \pi_{(1)c} \right\}^{b_c^{(j)}(1)} \right], \end{aligned} \quad (17)$$

where

$$b_c^{(j)}(1) = \begin{cases} 1 & \text{for } c > j \\ 0 & \text{for } c \leq j. \end{cases} \quad (18)$$

Next for the construction of the conditional likelihood at  $t$  given the information from previous time point  $t - 1$ , we first re-express the binary conditional probabilities in (12) and (14), as

$$\tilde{\lambda}_{gj}^{(2)}(g^*) = \begin{cases} \tilde{\lambda}_{gj}^{(2)}(1) & \text{for } b_{c_1}^{(g)}(t-1) = 0 \\ \tilde{\lambda}_{gj}^{(2)}(2) & \text{for } b_{c_1}^{(g)}(t-1) = 1, \end{cases} \quad (19)$$

$$\tilde{\lambda}_{gj}^{(1)}(g^*) = \begin{cases} \tilde{\lambda}_{gj}^{(1)}(1) & \text{for } b_{c_1}^{(g)}(t-1) = 0 \\ \tilde{\lambda}_{gj}^{(1)}(2) & \text{for } b_{c_1}^{(g)}(t-1) = 1. \end{cases} \quad (20)$$

One may then write the conditional likelihood for  $\beta$  and  $\gamma$ , as

$$L_{t|t-1}(\beta, \gamma) = \prod_{g=1}^{J-1} \prod_{j=1}^{J-1} \prod_{g^*=1}^2 \left[ \left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \right\}^{b_{c_2}^{(j)}(t)} \left\{ \tilde{\lambda}_{gj}^{(1)}(g^*) \right\}^{1-b_{c_2}^{(j)}(t)} \right], \quad (21)$$

where the binary data  $b_{c_2}^{(j)}(t)$  for observed  $c_2$  are obtained by (11), and similarly  $b_{c_1}^{(g)}(t - 1)$  to define  $g^*$  for given  $c_1$  are obtained from (10).

Next by combining (17) and (21), one obtains the likelihood function for  $\beta$  and  $\gamma$  as

$$\begin{aligned}
 L(\beta, \gamma) &= L_1(\beta) \prod_{t=2}^T L_{t|t-1}(\beta, \gamma) \\
 &= \prod_{j=1}^{J-1} \left[ \{F_{(1)j}\}^{1-b_{c_2}^{(j)}(1)} \right] \left[ \{1 - F_{(1)j}\}^{b_{c_2}^{(j)}(1)} \right] \\
 &\quad \times \prod_{t=2}^T \prod_{g=1}^{J-1} \prod_{j=1}^{J-1} \prod_{g^*=1}^2 \left[ \left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \right\}^{b_{c_2}^{(j)}(t)} \left\{ \tilde{\lambda}_{gj}^{(1)}(g^*) \right\}^{1-b_{c_2}^{(j)}(t)} \right]. \quad (22)
 \end{aligned}$$

For the benefit of the practitioners, we now develop the likelihood estimating equations for these parameters  $\beta$  and  $\gamma$ , as in the following sections. Remark that for the construction of similar likelihood estimating equations in the stationary longitudinal setup, one may be referred to Sutradhar [14, Sect. 3.6.2.2].

Note that the likelihood function in (22) is constructed by collapsing the ordinal multinomial responses to the binary responses at all suitable cut points. This likelihood function, therefore, can not be used for nominal multinomial time series data. When the categories are nominal, it is appropriate to construct the likelihood function by exploiting the marginal probability function  $\pi_{(t)j}$  from (2) for  $t = 1$ , and the conditional multinomial logit probability function  $\eta_{t|t-1}^{(j)}(g)$  from (4) for  $t = 2, \dots, T$  (see Loredo-Osti and Sutradhar [10]). Notice that in practice the time dependent covariates  $x_t$  in (2) and (4) are fixed in general. However, by treating  $x_t$  as a random covariate vector, Fokianos and Kedem [6] obtained parameter estimates by maximizing a partial likelihood function without requiring any extra characterization of the joint process  $\{y_t, x_t\}$ . Loredo-Osti and Sutradhar [10] have, however, argued that in Fokianos and Kedem’s [6] approach, the conditional Fisher information matrix is not the same as the one obtained by conditioning on  $\{x_t\}$ , the observed covariates. In fact, when the estimation is carried out in a general linear models framework that uses the canonical link function, this conditional information matrix obtained by Fokianos and Kedem, is just the Hessian matrix multiplied by  $-1$ , i.e., the observed information matrix.

As far as the ordinal multinomial time series data are concerned, the construction of binary mapping based likelihood function in (22) is a new concept. The core idea comes from the cumulative binary property for the MDL (multinomial dynamic logit) model (4) because of the present ordinal nature of the data. In the cross sectional setup, that is, for the case with  $t = 1$  only, the likelihood function for ordinal multinomial data has been used by many authors such as Agresti [1]. Note that the marginal multinomial probability in (2) has the multinomial logit form. In the cluster data setup, many existing studies use this multinomial logit model (2) as the marginal model at a given time  $t$ . As far as the correlations between repeated responses are concerned, some authors such as Agresti [1], Lipsitz et al. [9], Agresti and Natarajan [2] do not model them, rather they use ‘working’ correlations to construct the so-called generalized estimating equations and solve them to obtain the estimates for regression

parameters involved in the marginal multinomial logits model (2). These estimates however may not be reliable as they can be inefficient as compared to the ‘working’ independence assumption based estimates (see Sutradhar and Das [15], Sutradhar [13, Chap. 7] in the context of binary longitudinal data analysis). Thus, their extension to the time series setup may be useless. Moreover, it is not clear how to model the ordinal data using this type of ‘working’ correlations approach.

### 3.1 Likelihood Estimating Equations for the Regression Effects $\beta$

Recall that  $\beta = (\beta'_1, \dots, \beta'_j, \dots, \beta'_{j-1})' : (J - 1)(p + 1) \times 1$ , with  $\beta_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})'$ . For known  $\gamma$ , in this section, we exploit the likelihood function (22) and develop the likelihood estimating equation for  $\beta$ . For convenience, we use log likelihood function, which, following the likelihood function in (22), is written as

$$\begin{aligned} \text{Log } L(\beta, \gamma) &= \sum_{j=1}^{J-1} \left[ \{1 - b_c^{(j)}(1)\} \log F_{(1)j} + \{b_c^{(j)}(1)\} \log \{1 - F_{(1)j}\} \right] \\ &+ \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \left[ b_{c_2}^{(j)}(t) \log \left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \right\} + \{1 - b_{c_2}^{(j)}(t)\} \log \left\{ \tilde{\lambda}_{gj}^{(1)}(g^*) \right\} \right], \end{aligned} \tag{23}$$

yielding the likelihood estimating equation for  $\beta$  as

$$\begin{aligned} \frac{\partial \text{Log } L(\beta, \gamma)}{\partial \beta} &= \sum_{j=1}^{J-1} \left[ \frac{\{1 - b_c^{(j)}(1)\}}{F_{(1)j}} - \frac{\{b_c^{(j)}(1)\}}{\{1 - F_{(1)j}\}} \right] \frac{\partial F_{(1)j}}{\partial \beta} \\ &+ \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \left[ \frac{b_{c_2}^{(j)}(t)}{\tilde{\lambda}_{gj}^{(2)}(g^*)} - \frac{\{1 - b_{c_2}^{(j)}(t)\}}{\{1 - \tilde{\lambda}_{gj}^{(2)}(g^*)\}} \right] \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} \\ &= 0, \end{aligned} \tag{24}$$

where

$$\frac{\partial F_{(1)j}}{\partial \beta} = \sum_{c=1}^j \left[ \pi_{(1)c}(\delta_{(1)c} - \pi_{(1)}) \right] \otimes x_1; \tag{25}$$

and

$$\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} = \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=j+1}^J \left[ \eta_{t|t-1}^{(c_2)}(c_1)(\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1)) \right] \otimes x_t & \text{for } g^* = 1 \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=j+1}^J \left[ \eta_{t|t-1}^{(c_2)}(c_1)(\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1)) \right] \otimes x_t & \text{for } g^* = 2, \end{cases} \tag{26}$$



with

$$\begin{aligned} \pi_{(1)} &= [\pi_{(1)1}, \dots, \pi_{(1)c}, \dots, \pi_{(1)(J-1)}]' \\ \delta_{(t-1)c} &= \begin{cases} [01'_{c-1}, 1, 01'_{J-1-c}]' & \text{for } c = 1, \dots, J - 1 \\ 01_{J-1} & \text{for } c = J, \end{cases} \\ \eta_{t|t-1}(c_1) &= [\eta_{t|t-1}^{(1)}(c_1), \dots, \eta_{t|t-1}^{(c_2)}(c_1), \dots, \eta_{t|t-1}^{(J-1)}(c_1)]'. \end{aligned} \tag{27}$$

The details for the derivatives in (25) and (26) are given in ‘‘Appendix’’.

For given  $\gamma$ , the likelihood equations in (24) may be solved iteratively by using the iterative equation for  $\beta$  given by

$$\hat{\beta}(r+1) = \hat{\beta}(r) - \left[ \left\{ \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta' \partial \beta} \right\}^{-1} \frac{\partial \text{Log } L(\beta, \gamma)}{\partial \beta} \right]_{|\beta = \hat{\beta}(r)} ; (J-1)(p+1) \times 1, \tag{28}$$

where the formula for the second order derivative matrix  $\frac{\partial^2 \text{Log } L(\beta, \gamma_M)}{\partial \beta' \partial \beta}$  may be derived by taking the derivative of the  $(J-1)(p+1) \times 1$  vector with respect to  $\beta'$ . The exact second order derivative matrix has a complicated formula. We provide an approximation as follows.

**An approximation to  $\frac{\partial^2 \text{Log } L(\beta, \gamma_M)}{\partial \beta' \partial \beta}$  :**

Re-express the likelihood estimating equation from (24) as

$$\begin{aligned} \frac{\partial \text{Log } L(\beta, \gamma)}{\partial \beta} &= \sum_{j=1}^{J-1} \frac{\partial F_{(1)j}}{\partial \beta} \{ (1 - F_{(1)j}) F_{(1)j} \}^{-1} [ \{ 1 - b_c^{(j)}(1) \} - F_{(1)j} ] \\ + \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} &\left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \left( 1 - \tilde{\lambda}_{gj}^{(2)}(g^*) \right) \right\}^{-1} [ b_{c_2}^{(j)}(t) - \tilde{\lambda}_{gj}^{(2)}(g^*) ] \\ &= 0. \end{aligned} \tag{29}$$

Notice that in the first term in the left hand side of (29),  $\{ 1 - b_c^{(j)}(1) \}$  is, by (9), a binary variable with

$$\begin{aligned} E \{ 1 - b_c^{(j)}(1) \} &= F_{(1)j} \\ \text{var} \{ 1 - b_c^{(j)}(1) \} &= F_{(1)j} \{ 1 - F_{(1)j} \}, \end{aligned} \tag{30}$$

and similarly in the second term, by (12),  $b_{c_2}^{(j)}(t)$  conditional on  $b_{c_1}^{(g)}(t-1)$  is a binary variable with

$$\begin{aligned} E [ b_{c_2}^{(j)}(t) | b_{c_1}^{(g)}(t-1) ] &= \tilde{\lambda}_{gj}^{(2)}(g^*) \\ \text{var} [ b_{c_2}^{(j)}(t) | b_{c_1}^{(g)}(t-1) ] &= \tilde{\lambda}_{gj}^{(2)}(g^*) [ 1 - \tilde{\lambda}_{gj}^{(2)}(g^*) ], \end{aligned} \tag{31}$$

for  $g^* \equiv b_{c_1}^{(g)}(t - 1)$ . Thus, the likelihood estimating function in (29) is equivalent to a conditional quasi-likelihood (CQL) function in  $\beta$  for the cut points based binary data [e.g. see Tagore and Sutradhar [16, Eq. (27), p. 888]. Now because the variance of the binary data is a function of the mean, the variance and gradient functions in (29) may be treated to be known when mean is known. Thus, when a QL estimating equation is solved iteratively, the gradient and variance functions use  $\beta$  from a previous iteration [11, 18]. Consequently, by (29), the second derivative matrix required to compute (28) has a simpler approximate formula

$$\begin{aligned} \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \beta'} &= - \sum_{j=1}^{J-1} \frac{\partial F_{(1)j}}{\partial \beta} \{(1 - F_{(1)j})F_{(1)j}\}^{-1} \frac{\partial F_{(1)j}}{\partial \beta'} \\ &- \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} \left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \left(1 - \tilde{\lambda}_{gj}^{(2)}(g^*)\right) \right\}^{-1} \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta'}. \end{aligned} \quad (32)$$

Furthermore for known  $\gamma$ , by (28) and (29), under some mild conditions it follows that the solution of (29), say  $\hat{\beta}$ , satisfies

$$\hat{\beta} \sim N(\beta, V(\beta, \gamma)), \quad (33)$$

(see Kaufmann [8, Sect. 5]) where the covariance matrix is estimated by

$$\hat{V}(\cdot) = \left[ - \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \beta'} \right]_{\beta=\hat{\beta}}^{-1}. \quad (34)$$

### 3.2 Likelihood Estimating Equations for the Dynamic Dependence Parameters $\gamma$

In Sect. 3.1, we have estimated  $\beta$  for known  $\gamma$ , for example, initially by using  $\gamma = 0$ , where by (4)–(5),

$$\gamma = (\gamma'_1, \dots, \gamma'_j, \dots, \gamma'_{j-1})', \quad \text{with } \gamma_j = (\gamma_{j1}, \dots, \gamma_{jv}, \dots, \gamma_{j,J-1})'.$$

Note that  $F_{(1)j}$  for all  $j = 1, \dots, J - 1$ , are free from  $\gamma$ . Hence, by exploiting the log likelihood function (23), similar to (24), we write the likelihood equation for  $\gamma$  as

$$\frac{\partial \text{Log } L(\beta, \gamma)}{\partial \gamma} = \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \left[ \frac{b_{c_2}^{(j)}(t)}{\tilde{\lambda}_{gj}^{(2)}(g^*)} - \frac{\{1 - b_{c_2}^{(j)}(t)\}}{\{1 - \tilde{\lambda}_{gj}^{(2)}(g^*)\}} \right] \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma} = 0, \quad (35)$$

where

$$\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma} = \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=j+1}^J \left[ \eta_{t|t-1}^{(c_2)}(c_1) (\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1)) \right] \otimes \delta_{(t-1)c_1} & \text{for } g^* = 1 \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=j+1}^J \left[ \eta_{t|t-1}^{(c_2)}(c_1) (\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1)) \right] \otimes \delta_{(t-1)c_1} & \text{for } g^* = 2. \end{cases} \quad (36)$$

An outline for this derivative is given in the ‘‘Appendix’’.

By similar calculations as in (28), one may solve the likelihood estimating equation in (35) for  $\gamma$  using the iterative equation

$$\hat{\gamma}(r+1) = \hat{\gamma}(r) - \left[ \left\{ \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \gamma'} \right\}^{-1} \frac{\partial \text{Log } L(\beta, \gamma)}{\partial \gamma} \right]_{|\gamma=\hat{\gamma}(r)} ; (J-1)^2 \times 1, \quad (37)$$

where the second order derivative matrix, following (32), may be computed as

$$\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \gamma'} = - \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma} \left\{ \tilde{\lambda}_{gj}^{(2)}(g^*) \left( 1 - \tilde{\lambda}_{gj}^{(2)}(g^*) \right) \right\}^{-1} \frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma'}. \quad (38)$$

Furthermore for known  $\beta$ , by (37) and (38), it follows under some mild conditions that the solution of (35), say  $\hat{\gamma}$ , satisfies

$$\hat{\gamma} \sim N(\gamma, V^*(\beta, \gamma)), \quad (39)$$

(see Kaufmann [8, Sect. 5]) where the covariance matrix is estimated by

$$\hat{V}^*(\cdot) = \left[ - \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \gamma'} \right]_{|\gamma=\hat{\gamma}}^{-1}. \quad (40)$$

### 3.3 Joint Likelihood Estimating Equations for $\beta$ and $\gamma$

Let  $\theta = (\beta', \gamma')'$ . One may then combine (28) and (37) and solve the iterative equation

$$\hat{\theta}(r+1) = \hat{\theta}(r) - \left[ \left( \frac{\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \beta'}}{\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \beta'}} \quad \frac{\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \gamma'}}{\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \gamma'}} \right)^{-1} \left( \frac{\frac{\partial \text{Log } L(\beta, \gamma)}{\partial \beta}}{\frac{\partial \text{Log } L(\beta, \gamma)}{\partial \gamma}} \right) \right]_{|\theta=\hat{\theta}(r)} \quad (41)$$

to obtain the joint likelihood estimates for  $\beta$  and  $\gamma$ . In order to construct the iterative equation (41), we require the formula for the second order derivative matrix  $\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \gamma'}$  which, using (29), may be approximately computed as

$$\frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \gamma'} = - \sum_{t=2}^T \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \sum_{g^*=1}^2 \frac{\partial \bar{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} \left\{ \bar{\lambda}_{gj}^{(2)}(g^*) \left( 1 - \bar{\lambda}_{gj}^{(2)}(g^*) \right) \right\}^{-1} \frac{\partial \bar{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma'}, \quad (42)$$

where the formulas for  $\frac{\partial \bar{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta}$  and  $\frac{\partial \bar{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma}$  are given by (26) and (36), respectively.

Furthermore, by similar arguments to (33) and (39), under some mild conditions it follows that the solution of (41), say  $\begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix}$  has the multivariate Gaussian distribution

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix} \sim N \left[ \begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} \tilde{V}_{11}(\beta, \gamma) & \tilde{V}_{12}(\beta, \gamma) \\ \tilde{V}'_{12}(\beta, \gamma) & \tilde{V}_{22}(\beta, \gamma) \end{pmatrix} \right], \quad (43)$$

where  $\text{cov}(\tilde{\beta}) = \tilde{V}_{11}(\beta, \gamma)$  and  $\text{cov}(\tilde{\gamma}) = \tilde{V}_{22}(\beta, \gamma)$  are estimated as

$$\begin{aligned} \widehat{\text{cov}}(\tilde{\beta}) &= A^{-1} + FE^{-1}F' \\ \widehat{\text{cov}}(\tilde{\gamma}) &= E^{-1}, \end{aligned} \quad (44)$$

Rao [12, p. 33] with  $E = D - B'A^{-1}B$ , and  $F = A^{-1}B$ , where by (41)

$$A = \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \beta'}; \quad B = \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \beta \partial \gamma'}; \quad \text{and} \quad D = \frac{\partial^2 \text{Log } L(\beta, \gamma)}{\partial \gamma \partial \gamma'}.$$

## 4 Concluding Remarks

Recently some authors such as Loredo-Osti and Sutradhar [10] (see also Fokianos and Kedem [6]) have developed a likelihood approach for the estimation of regression and dynamic dependence parameters involved in a multinomial dynamic logit (MDL) model used for categorical time series data. This inference issue becomes more complex when the categorical response collected at a given time point also exhibit an order. In this paper we have demonstrated that this type of ordinal categorical responses collected over time may be analyzed by collapsing a multinomial response to a binary response at a given possible cut point and fitting binary dynamic model to all such binary responses collected based on all possible cut points over all times. For simplicity, we have fitted a low order, namely lag 1 dynamic model among all possible cut points based binary responses. A pseudo likelihood method using binary responses (in stead of the multinomial observations) is then constructed for the estimation of the regression and dynamic dependence parameters. The authors plan to undertake an empirical study involving simulations and real life data analysis in order to investigate the performance of the proposed estimation approach both for moderate and large size time series. The empirical results will be published elsewhere.

**Acknowledgments** The authors are grateful to Bhagawan Sri Sathya Sai Baba for His love and blessings to carry out this research in Sri Sathya Institute of Higher Learning. The authors thank the editorial committee for the invitation to participate in preparing this Festschrift honoring Professor Ian McLeod. It has brought back many pleasant memories of Western in early 80's experienced by the first author during his PhD study. We have prepared this small contribution as a token of our love and respect to Professor Ian McLeod for his long and sustained contributions to the statistics community through teaching and research in time series analysis, among other areas. The authors thank two referees for their comments and suggestions on the earlier version of the paper.

## Appendix

*Derivation for  $\frac{\partial F_{(1)j}}{\partial \beta}$  :*

Recall from Sect. 2.1 that  $F_{(1)j} = \sum_{c=1}^j \pi_{(1)c}$ , where  $\pi_{(1)c}$  is given by (2). It then follows that

$$\frac{\partial F_{(1)j}}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{c=1}^j \pi_{(1)c} = \frac{\partial}{\partial \beta} \sum_{c=1}^j \frac{\exp(x'_1 \beta_c)}{1 + \sum_{g=1}^{J-1} \exp(x'_1 \beta_g)}. \tag{45}$$

Now because

$$\frac{\partial \pi_{(1)c}}{\partial \beta_c} = \pi_{(1)c}[1 - \pi_{(1)c}]x_1, \text{ and } \frac{\partial \pi_{(1)c}}{\partial \beta_k} = -[\pi_{(1)c}\pi_{(1)k}]x_1, \tag{46}$$

it follows that

$$\begin{aligned} \frac{\partial \pi_{(1)c}}{\partial \beta} &= \begin{pmatrix} -\pi_{(1)1}\pi_{(1)c} \\ \vdots \\ \pi_{(1)c}[1 - \pi_{(1)c}] \\ \vdots \\ -\pi_{(1)(J-1)}\pi_{(1)c} \end{pmatrix} \otimes x_1 : (J - 1)(p + 1) \times 1 \\ &= [\pi_{(1)c}(\delta_{(1)c} - \pi_{(1)})] \otimes x_1. \end{aligned} \tag{47}$$

The formula for  $\frac{\partial F_{(1)j}}{\partial \beta}$  in (25) follows by using (47) and (45).

*Derivation for  $\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta}$  :*

By using the formula for  $\tilde{\lambda}_{gj}^{(2)}(g^*)$  from (13) we write

$$\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \beta} = \frac{\partial}{\partial \beta} \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) & \text{for } g^* = 1 \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) & \text{for } g^* = 2, \end{cases} \tag{48}$$

where  $\lambda_{t|t-1}^{(c_2)}(c_1)$  is given in (5), that is,

$$\eta_{t|t-1}^{(c_2)}(c_1) = \begin{cases} \frac{\exp[x'_t \beta_{c_2} + \gamma'_{c_2} \delta_{(t-1)c_1}]}{1 + \sum_{v=1}^{J-1} \exp[x'_t \beta_v + \gamma'_v \delta_{(t-1)c_1}]}, & \text{for } c_2 = 1, \dots, J-1 \\ \frac{1}{1 + \sum_{v=1}^{J-1} \exp[x'_t \beta_v + \gamma'_v \delta_{(t-1)c_1}]}, & \text{for } c_2 = J. \end{cases} \quad (49)$$

Now, for  $t = 2, \dots, T$ , it follows from (49) that

$$\begin{aligned} \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \beta_{c_2}} &= \eta_{t|t-1}^{(c_2)}(c_1) [1 - \eta_{t|t-1}^{(c_2)}(c_1)] x_t \\ \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \beta_k} &= - [\eta_{t|t-1}^{(c_2)}(c_1) \eta_{t|t-1}^{(k)}(c_1)] x_t, \end{aligned} \quad (50)$$

yielding

$$\begin{aligned} \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \beta} &= \begin{pmatrix} -\eta_{t|t-1}^{(1)}(c_1) \eta_{t|t-1}^{(c_2)}(c_1) \\ \vdots \\ \eta_{t|t-1}^{(c_2)}(c_1) [1 - \eta_{t|t-1}^{(c_2)}(c_1)] \\ \vdots \\ -\eta_{t|t-1}^{(J-1)}(c_1) \eta_{t|t-1}^{(c_2)}(c_1) \end{pmatrix} \otimes x_t : (J-1)(p+1) \times 1 \\ &= [\eta_{t|t-1}^{(c_2)}(c_1) (\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1))] \otimes x_t. \end{aligned} \quad (51)$$

The formula for the derivative in (26) follows now by applying (50) into (48).

*Derivation for  $\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma}$ :*

By using the formula for  $\tilde{\lambda}_{gj}^{(2)}(g^*)$  from (13) we write

$$\frac{\partial \tilde{\lambda}_{gj}^{(2)}(g^*)}{\partial \gamma} = \frac{\partial}{\partial \gamma} \begin{cases} \frac{1}{g} \sum_{c_1=1}^g \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) & \text{for } g^* = 1 \\ \frac{1}{J-g} \sum_{c_1=g+1}^J \sum_{c_2=j+1}^J \lambda_{t|t-1}^{(c_2)}(c_1) & \text{for } g^* = 2, \end{cases} \quad (52)$$

where  $\lambda_{t|t-1}^{(c_2)}(c_1)$  is given in (5) [see also (49)].

Next, for  $t = 2, \dots, T$ , it follows from (49) that

$$\begin{aligned} \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \gamma_{c_2}} &= \eta_{t|t-1}^{(c_2)}(c_1) [1 - \eta_{t|t-1}^{(c_2)}(c_1)] \delta_{(t-1)c_1} \\ \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \gamma_k} &= - [\eta_{t|t-1}^{(c_2)}(c_1) \eta_{t|t-1}^{(k)}(c_1)] \delta_{(t-1)c_1}, \end{aligned} \quad (53)$$

where

$$\delta_{(t-1)c_1} = \begin{cases} [01'_{c_1-1}, 1, 01'_{J-1-c_1}]' & \text{for } c_1 = 1, \dots, J-1 \\ 01_{J-1} & \text{for } c_1 = J. \end{cases}$$

These derivatives in (53) may further be re-expressed as

$$\begin{aligned} \frac{\partial \eta_{t|t-1}^{(c_2)}(c_1)}{\partial \gamma} &= \begin{pmatrix} -\eta_{t|t-1}^{(1)}(c_1)\eta_{t|t-1}^{(c_2)}(c_1) \\ \vdots \\ \eta_{t|t-1}^{(c_2)}(c_1)[1 - \eta_{t|t-1}^{(c_2)}(c_1)] \\ \vdots \\ -\eta_{t|t-1}^{(J-1)}(c_1)\eta_{t|t-1}^{(c_2)}(c_1) \end{pmatrix} \otimes \delta_{(t-1)c_1} : (J-1)^2 \times 1 \\ &= \left[ \eta_{t|t-1}^{(c_2)}(c_1)(\delta_{(t-1)c_2} - \eta_{t|t-1}(c_1)) \right] \otimes \delta_{(t-1)c_1}. \end{aligned} \tag{54}$$

The formula for the derivative in (36) now follows by applying (53) into (52).

## References

1. Agresti, A. (1989). A survey of models for repeated ordered categorical response data. *Statistics in Medicine*, 8, 1209–1224.
2. Agresti, A., & Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review*, 69, 345–371.
3. Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.
4. Fahrmeir, L., & Kaufmann, H. (1987). Regression models for non-stationary categorical time series. *Journal of Time Series Analysis*, 8, 147–160.
5. Fokianos, K., & Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis*, 67, 277–296.
6. Fokianos, K., & Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science*, 18, 357–376.
7. Fokianos, K., & Kedem, B. (2004). Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, 25, 173–197.
8. Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Annals of Statistics*, 15, 79–98.
9. Lipsitz, S. R., Kim, K., & Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13, 1149–1163.
10. Loredo-Osti, J. C., & Sutradhar, B. C. (2012). Estimation of regression and dynamic dependence parameters for non-stationary multinomial time series. *Journal of Time Series Analysis*, 33, 458–467.
11. McCullagh, P. (1983). Quasilikelihood functions. *Annals of Statistics*, 11, 59–67.
12. Rao, C. R. (1973). *Linear statistical inference and its applications*. New York, NY: Wiley.
13. Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data*. New York, NY: Springer.
14. Sutradhar, B. C. (2014). *Longitudinal categorical data analysis*. New York, NY: Springer.
15. Sutradhar, B. C., & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, 86, 459–465.
16. Tagore, V., & Sutradhar, B. C. (2009). Conditional inference in linear versus nonlinear models for binary time series. *Journal of Statistical Computation and Simulation*, 79, 881–897.

17. Tong, H. (1990). *Nonlinear time series: A dynamical system approach*. Oxford statistical science series (Vol. 6). New York, NY: Oxford University Press (1990)
18. Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61, 439–447.