

TECHNICAL RESEARCH REPORT

Partial Likelihood Analysis of Categorical Time Series Models

*by K. Fokianos, B. Kedem, and
D.A. Short*

T.R. 95-86



*Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry*

Partial Likelihood Analysis of Categorical Time Series Models

by

Konstantinos Fokianos, Benjamin Kedem*, and David A. Short**

*Mathematics Department and Institute for Systems Research
University of Maryland, College Park
Maryland 20742, USA

and

**Laboratory of Atmospheres
NASA/GSFC, Greenbelt
Maryland 20771, USA

September 1995 ¹

¹Work supported by NSF grant EEC-940-2384, and NASA contract NAG52783

Abstract

Partial likelihood analysis of two generalized logistic regression models for nominal and ordinal categorical time series is presented, taking into account stochastic time dependent covariates. Under some conditions on the covariates, the resulting estimators are consistent and asymptotically normal. The analysis is applied to rainfall data where the goodness of fit is judged by a certain chi square statistic.

Keywords: Time dependent covariates, ordinal, autoregression, nonstationary, martingale, goodness of fit.

1 Introduction

With the advent of generalized linear models—described in McCullagh and Nelder (1989)—there is much recent interest in categorical time series and their application. The recent books by Diggle, Liang and Zeger (1994), Fahrmeir and Tutz (1994), and Kedem (1994, Ch. 9), all attest to this trend. Categorical time series arise naturally when observing nominal and ordinal time records, or when a time series undergoes quantization at several levels.

A particular example of the latter type is the classification of precipitation radar data into bins in the effective dynamic range as explained in Meneghini and Jones (1993). We shall consider a similar case below using a time series of rain rate averaged over an area every 10 minutes. As we shall see, an important covariate for this problem is the fraction of the area where it is raining.

The objective of this paper is to discuss partial likelihood estimation and a goodness of fit procedure for ordinal and nominal categorical time series models, following the logistic regression paradigm. Partial likelihood—introduced by Cox (1975), extended and ramified in Wong (1986), Slud (1992), and Andersen, Borgan, Gill and Keiding (1993)—generalizes both the notion of likelihood and conditional likelihood, and is particularly useful for time series data where the precise dependence structure is not known.

We mainly focus on two models, the *multinomial logits model* and the *proportional odds model*. In both cases, under some conditions the maximum partial likelihood estimator (MPLE) is consistent and asymptotically normal—a fact established by appealing to martingale theory.

2 Partial Likelihood Considerations

Assume that an individual observes a stochastic process, say (x_t, y_t) , $t = 1, \dots, N$. In principle, we can write down the joint distribution of all the observations up to time N , by employing the law of total probability; that is (Wong (1986))

$$f(x_1, y_1, x_2, y_2, \dots, x_N, y_N) = \left[\prod_{t=1}^N f(y_t \mid d_t) \right] \left[\prod_{t=1}^N f(x_t \mid c_t) \right] \quad (2.1)$$

where $d_t = (y_1, x_1, \dots, y_{t-1}, x_{t-1})$ and $c_t = (y_1, x_1, \dots, y_{t-1}, x_{t-1}, y_t)$.

Cox (1975) defined the second product on the right hand side of (2.1) as the Partial Likelihood. It is helpful to note that the σ -field generated by c_{t-1} is contained in the one

generated by c_t . This is a key feature which motivates our definition (see Slud (1992), and Slud and Kedem (1994)).

Definition 2.1 Let $\mathcal{F}_t, t = 0, 1, \dots$ be an increasing sequence of σ -fields, and let X_1, X_2, \dots be a sequence of random variables in some common probability space such that X_t is \mathcal{F}_t measurable. Denote the density of X_t given \mathcal{F}_{t-1} by $f_t(x_t; \beta)$, where $\beta \in R^p$ is a parameter. The *Partial Likelihood* function relative to β, \mathcal{F}_t , and the data X_1, X_2, \dots, X_N , is given by the product

$$PL(\beta; X_1, \dots, X_N) = \prod_{t=1}^N f_t(x_t; \beta) \quad (2.2)$$

This definition generalizes both likelihood and conditional likelihood. Unlike (full) likelihood, partial likelihood does not require complete knowledge of the joint distribution of the covariates. Unlike conditional likelihood, complete covariate information need not be known throughout the period of observation. Partial likelihood takes into account only what is known to the observer up to the time of actual observation.

The vector $\hat{\beta}$ that maximizes (2.2) is called the maximum partial likelihood estimator (MPLE). Its asymptotic distribution has been studied by several authors (see Wong (1986) ; Slud and Kedem (1994)). The key point is that the gradient of the logarithm of (2.2) is a martingale with respect to the nested sequence of histories \mathcal{F}_t .

Before we proceed, we need to establish some notation and make some calculations which will be found useful in the sequel.

2.1 The Mathematical Setup

Assume that we observe a nonstationary time series, say $\{Y_t\}$, with m possible categories for each observation. Let \mathbf{z}_t denote a vector of random time dependent covariates. This may contain lagged values of the observations process or any other time series which evolves in time simultaneously with Y_t . Suppose that the t 'th observation is given by a vector $\mathbf{y}_t = (y_{t1}, \dots, y_{tq})'$ of length $q = m - 1$, where

$$y_{tj} = \begin{cases} 1 & \text{if the } j\text{'th category is observed at time } t \\ 0 & \text{otherwise} \end{cases}$$

Denote by $\mathbf{p}_t = (p_{t1}, \dots, p_{tq})'$ the corresponding vector of conditional probabilities given \mathcal{F}_{t-1} , that is $p_{tj} = P(y_{tj} = 1 \mid \mathcal{F}_{t-1})$, $j = 1, \dots, q$, and \mathcal{F}_{t-1} stands for the available

information to the observer up to and including time t . For the m' th category, put

$$y_{tm} = 1 - \sum_{j=1}^q y_{tj} \quad (2.3)$$

and

$$p_{tm} = 1 - \sum_{j=1}^q p_{tj} \quad (2.4)$$

By parametrizing suitably the probabilities of each category, the partial likelihood based on the sample $\mathbf{y}_1, \dots, \mathbf{y}_N$, is easily obtained. Since each component of \mathbf{y}_t takes the values 0 or 1 we have the multinomial probability

$$f(y_t; \boldsymbol{\beta} \mid \mathcal{F}_{t-1}) = \prod_{j=1}^m p_{tj}(\boldsymbol{\beta})^{y_{tj}} \quad (2.5)$$

Consequently, the corresponding Partial Likelihood is:

$$\begin{aligned} PL(\boldsymbol{\beta}) &= \prod_{t=1}^N f(y_t; \boldsymbol{\beta} \mid \mathcal{F}_{t-1}) \\ &= \prod_{t=1}^N \prod_{j=1}^m p_{tj}(\boldsymbol{\beta})^{y_{tj}} \end{aligned} \quad (2.6)$$

It follows that the partial log-likelihood is given by

$$pl_N(\boldsymbol{\beta}) = \sum_{t=1}^N \sum_{j=1}^m y_{tj} \log p_{tj}(\boldsymbol{\beta}) \quad (2.7)$$

The partial score is given by the vector

$$S_N(\boldsymbol{\beta}) = \left(\frac{\partial pl_N(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial pl_N(\boldsymbol{\beta})}{\partial \beta_p} \right)' \quad (2.8)$$

The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$ is a consistent root of the equation $S_N(\boldsymbol{\beta}) = 0$. As in full likelihood inference, we need the notion of *information matrix*, given here by the conditional information matrix

$$\mathbf{G}_N(\boldsymbol{\beta}) = \sum_{t=1}^N \text{Var}(\boldsymbol{\alpha}_t(\boldsymbol{\beta}) \mid \mathcal{F}_{t-1}) \quad (2.9)$$

with $\alpha_t(\beta) = S_t(\beta) - S_{t-1}(\beta)$. The unconditional information matrix is given by

$$\mathbf{F}_N(\beta) = E[\mathbf{G}_N(\beta)] \quad (2.10)$$

Finally, let $\mathbf{H}_N(\beta)$ denote the negative matrix of the second partial derivatives of the partial log-likelihood, that is

$$\mathbf{H}_N(\beta) = -\frac{\partial^2 p l_N(\beta)}{\partial \beta \partial \beta'}$$

Notice, that the expectation and variance have been taken above with respect to the true parameter β_0 .

3 Multinomial Logits Model

When y_t is binary, logistic regression with time dependent covariates is defined by the model (Korn and Whittemore (1979); Zeger et al. (1985) ; Cox and Snell (1989); Slud and Kedem (1994))

$$P(y_t = 1 \mid \mathcal{F}_{t-1}) = \frac{1}{1 + \exp(-\beta' \mathbf{z}_{t-1})} \quad (3.1)$$

where \mathbf{z}_{t-1} may contain past values of y_t . It has been shown (Slud and Kedem (1994)) that the MPLE exists, is consistent and asymptotically normal. A generalization aimed at nominal time series is as follows (Agresti (1990)).

Assume that we observe a multicategorical time series as in section 2.1. Then in analogy with (3.1) put,

$$\log \frac{p_{tj}}{p_{tm}} = \beta'_j \mathbf{z}_{t-1}, \quad (j = 1, \dots, q) \quad (3.2)$$

Recall that $q = m - 1$. It follows from (2.4), that

$$p_{tj} = \frac{\exp(\beta'_j \mathbf{z}_{t-1})}{1 + \sum_{i=1}^q \exp(\beta'_i \mathbf{z}_{t-1})}, \quad (j = 1, \dots, q) \quad (3.3)$$

where β_j is a p -dimensional regression parameter and \mathbf{z}_{t-1} is a vector of stochastic time dependent covariates of the same dimension. Another derivation of this model is described in Mc Fadden (1973) by maximizing a utility function.

Observe from (3.3) that

$$\log \frac{p_{tj}}{p_{ti}} = (\beta'_j - \beta'_i) \mathbf{z}_{t-1}$$

So, we see that the ratio, p_{tj}/p_{ti} , for the j 'th and i 'th category is the same irrespective of the total number of categories m . This property is usually referred as *independence of irrelevant alternatives*. The function

$$l(\mathbf{p}_t) = (\log \frac{p_{t1}}{p_{tm}}, \dots, \log \frac{p_{tq}}{p_{tm}}) \quad (3.4)$$

is called the *logit* function.

In this section we let $\boldsymbol{\beta}$ be the pq -vector

$$\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_q)'$$

and \mathbf{Z}_{t-1} be the $qp \times q$ matrix

$$\mathbf{Z}_{t-1} = \begin{bmatrix} \mathbf{z}_{t-1} & 0 & \cdots & 0 \\ 0 & \mathbf{z}_{t-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{z}_{t-1} \end{bmatrix}$$

The partial score function is easily obtained by means of properties of the exponential family as (McCullagh and Nelder (1989, ch. 2); Kedem (1994, pp. 288-290))

$$S_N(\boldsymbol{\beta}) = \sum_{t=1}^N \mathbf{Z}_{t-1}(\mathbf{y}_t - \mathbf{p}_t(\boldsymbol{\beta})) \quad (3.5)$$

It readily follows that

$$\mathbf{G}_N(\boldsymbol{\beta}) = \sum_{t=1}^N \mathbf{Z}_{t-1} \boldsymbol{\Sigma}_t(\boldsymbol{\beta}) \mathbf{Z}_{t-1}'$$

with the (i, j) element of $\boldsymbol{\Sigma}_t(\boldsymbol{\beta})$ is given by

$$\sigma_t^{(ij)}(\boldsymbol{\beta}) = \begin{cases} -p_{ti}(\boldsymbol{\beta})p_{tj}(\boldsymbol{\beta}) & \text{if } i \neq j \\ p_{ti}(\boldsymbol{\beta})(1 - p_{ti}(\boldsymbol{\beta})) & \text{if } i = j \end{cases}$$

for $i, j = 1, \dots, q$. It also follows that the negative matrix of the second partial derivatives, that is the sample information matrix, satisfies the equality

$$\mathbf{H}_N(\boldsymbol{\beta}) = \mathbf{G}_N(\boldsymbol{\beta})$$

The last relation—a consequence of the multinomial logits model (3.3)—is of great importance, since it implies concavity of the partial log-likelihood and therefore uniqueness of the corresponding estimator.

3.1 Large Sample Theory

We will prove now existence, consistency and asymptotic Normality of the MPLE under regularity conditions.

Assumption (A)

- A.1** The parameter β belongs to an open set $B \subseteq R^{pq}$.
- A.2** The covariate matrix belongs to a non-random compact subset of $R^{pq \times q}$ such that $\sum_{t=1}^N \mathbf{Z}_{t-1} \mathbf{Z}'_{t-1}$ is positive definite almost everywhere.
- A.3** The probability measure P which governs $\{\mathbf{y}_t, \mathbf{Z}_t\}$, $t = 1, \dots, N$ gives (3.3) with $\beta = \beta_0$.
- A.4** There is a probability measure ν on $R^{pq \times q}$ such that $\int_{R^{pq \times q}} \mathbf{Z} \mathbf{Z}' d\nu(\mathbf{Z})$ is positive definite, such that under (3.3) with $\beta = \beta_0$, and for Borel sets A we have

$$\frac{1}{N} \sum_{t=1}^N I_{[\mathbf{Z}_{t-1} \in A]} \xrightarrow{P} \nu(A), \quad N \rightarrow \infty \quad (3.6)$$

Assumption **A.2** is useful for the derivation of bounds for asymptotics. Furthermore, it ensures that the conditional information matrix is positive definite with probability 1. It follows that the unconditional information matrix is positive definite. Due to assumption **A.4**, we can conclude that the conditional information matrix has a limit in probability

$$\frac{\mathbf{G}_N(\beta)}{N} \xrightarrow{P} \int_{R^{pq \times q}} \mathbf{Z} \Sigma(\beta) \mathbf{Z}' d\nu(\mathbf{Z}) \equiv \Lambda_1(\beta) \quad (3.7)$$

where Σ has the generic element

$$\sigma^{(ij)}(\beta) = \begin{cases} -p_i(\mathbf{Z}'\beta)p_j(\mathbf{Z}'\beta) & \text{if } i \neq j \\ p_i(\mathbf{Z}'\beta)(1 - p_i(\mathbf{Z}'\beta)) & \text{if } i = j \end{cases}$$

for $i, j = 1, \dots, q$.

Here, integration with respect to a matrix means that we integrate with respect to each element of the matrix. It follows from **A.4** that $\Lambda_1(\beta)$ is a positive definite matrix and therefore its inverse exist at the true value. It is important to emphasize that our approach is quite general and does not call for any Markov assumption(compare with Fahrmeir and Kaufmann (1987); Kaufmann (1987)).

The main result of this section is the following theorem.

Theorem 3.1 Consider the multinomial logits model (3.3) and assume that assumption (A) holds. Then we have that:

1. There exists a unique MPLE, $\hat{\beta}$, with probability tending to 1, as $N \rightarrow \infty$.
2. The estimator is consistent and asymptotically normal,

$$\hat{\beta} \xrightarrow{p} \beta_0$$

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{D} \mathcal{N}(0, \Lambda_1^{-1}(\beta_0))$$

Proof: We sketch the proof of the theorem. Since all the calculations are under the true value we drop the dependence of the notation on β_0 . Notice that the partial score (3.5) is a qp -variate zero mean square integrable martingale with respect to $\{\mathcal{F}_N\}$. Appealing to the Cramer-Wold device (Billingsley (1986, Ch. 29)), we define $\phi_N = \lambda' S_N$ for some $\lambda \in R^{pq}$. Then ϕ_N is a univariate zero mean square integrable martingale. Its conditional and unconditional covariance matrices are $\lambda' \mathbf{G}_N \lambda$ and $\lambda' \mathbf{F}_N \lambda$ respectively. We have,

$$\frac{\lambda' \mathbf{G}_N \lambda}{\lambda' \mathbf{F}_N \lambda} \xrightarrow{p} 1$$

and

$$\frac{1}{\lambda' \mathbf{F}_N \lambda} \sum_{t=1}^N E[|a_t|^2 I_{Nt}(\epsilon) \mid \mathcal{F}_{t-1}] \xrightarrow{p} 0$$

as $N \rightarrow \infty$, with $I_{Nt}(\epsilon) = I_{[|a_t|^2 \geq (\lambda' \mathbf{F}_N \lambda)^{1/2} \epsilon]}$ and $a_t = \phi_t - \phi_{t-1}$. It follows from the Central Limit Theorem for martingales (Hall and Heyde (1980, Corollary 3.1)) that

$$\frac{S_N}{\sqrt{N}} \xrightarrow{D} \mathcal{N}(0, \Lambda_1)$$

Asymptotic existence can be established along the lines of Kaufmann (1987) by proving that for every $\eta > 0$ there exists N and δ such that

$$P[pl_N(\beta) - pl_N(\beta_0) < 0 \ \forall \beta \in \partial O_N(\delta)] \geq 1 - \eta \quad (3.8)$$

with $O_N(\delta) = \{\beta : \|\mathbf{F}_N^{1/2}(\beta - \beta_0)\| \leq \delta\}$. The basic idea is to use a Taylor expansion of the log-partial likelihood and estimate the above probability. Since for the multinomial

logits model $\mathbf{H}_N = \mathbf{G}_N$, it follows that the MPLE $\hat{\beta}$ is also unique with probability tending to 1. From the above discussion, we immediately get that

$$\begin{aligned} 1 - \eta &\leq P(\|\mathbf{F}_N^{1/2}(\hat{\beta} - \beta_0)\| \leq \delta) \\ &\leq P(\|\hat{\beta} - \beta_0\| \leq \epsilon) \end{aligned}$$

for some ϵ . Therefore the estimator is also consistent. The last step makes use of the mean value theorem

$$S_N = \mathbf{G}_N(\tilde{\beta})(\hat{\beta} - \beta_0)$$

for some $\tilde{\beta}$ in the line segment connecting $\hat{\beta}$ and β_0 . Since \mathbf{G}_N is a continuous function of β , from the consistency that was just proved we have

$$\mathbf{G}_N(\tilde{\beta})(\hat{\beta} - \beta_0) \approx \mathbf{G}_N(\hat{\beta})(\hat{\beta} - \beta_0)$$

The desired result follows from the Central Limit Theorem we proved earlier. \square

4 The Proportional Odds Model

4.1 On the model

One of the most widely used models for the analysis of ordinal data is the proportional odds model (Snell (1964); McCullagh (1980)). We show how one can derive this model by using the method of a latent variable.

Assume that $x_t = -\gamma' \mathbf{z}_{t-1} + e_t$, where e_t is a sequence of i.i.d logistically distributed random variables, γ is a vector of parameters and \mathbf{z}_{t-1} is a covariate vector of the same dimension. Suppose that we observe

$$y_t = j \iff \theta_{j-1} \leq x_t < \theta_j$$

for $j = 1, \dots, q$, where $-\infty = \theta_0 < \theta_1 < \dots < \theta_k = \infty$ are the so called threshold parameters. It follows that

$$\begin{aligned} P(y_t = j \mid \mathcal{F}_{t-1}) &= P(\theta_{j-1} \leq x_t < \theta_j \mid \mathcal{F}_{t-1}) \\ &= F(\theta_j + \gamma' \mathbf{z}_{t-1}) - F(\theta_{j-1} + \gamma' \mathbf{z}_{t-1}) \end{aligned}$$

with $F(x) = 1/(1 + \exp(-x))$. The model can be formulated somewhat more compactly by the equation:

$$P(y_t \leq j \mid \mathcal{F}_{t-1}) = F(\theta_j + \gamma' \mathbf{z}_{t-1}) \tag{4.1}$$

Since the set of cumulative probabilities corresponds one to one to the set of the response probabilities, estimating the former enables estimation of the latter.

In this section we let β to be the $q + p$ vector

$$\beta = (\theta_1, \dots, \theta_q, \gamma')'$$

and

$$\mathbf{Z}'_{t-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \mathbf{z}_{t-1} \\ 0 & 1 & \cdots & 0 & \mathbf{z}_{t-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \mathbf{z}_{t-1} \end{bmatrix}$$

$$h = (h_1, \dots, h_q)$$

with

$$\begin{aligned} h_1(\beta) &= F(\eta_{(t-1)1}(\beta)) \\ h_j(\beta) &= F(\eta_{(t-1)j}(\beta)) - F(\eta_{(t-1)(j-1)}(\beta)), \quad j = 2, \dots, q \end{aligned}$$

and $h_m = 1 - \sum_{j=1}^q h_j$ where

$$\boldsymbol{\eta}_{t-1} = (\eta_{(t-1)1}, \dots, \eta_{(t-1)q}) = \mathbf{Z}_{t-1}\beta$$

With this notation, the partial score function for the proportional odds model becomes

$$S_N(\beta) = \sum_{t=1}^N \mathbf{Z}_{t-1} \mathbf{U}_{t-1}(\beta) (\mathbf{y}_t - \mathbf{p}_t(\beta)) \quad (4.2)$$

with $\mathbf{U}_{t-1}(\beta) = [\partial(l \circ h)/\partial \boldsymbol{\eta}_{t-1}]$ where l is the logit function (3.4). It follows that the conditional information matrix is

$$\mathbf{G}_N(\beta) = \sum_{t=1}^N \mathbf{Z}_{t-1} \mathbf{U}_{t-1}(\beta) \boldsymbol{\Sigma}_t(\beta) \mathbf{U}'_{t-1}(\beta) \mathbf{Z}'_{t-1}$$

with $\boldsymbol{\Sigma}_t(\beta)$ is as before. The negative matrix of the partial second derivatives is

$$\mathbf{H}_N(\beta) = \mathbf{G}_N(\beta) - \mathbf{R}_N(\beta) \quad (4.3)$$

with $\mathbf{R}_N(\beta) = \sum_{t=1}^N \sum_{r=1}^q \mathbf{Z}_{t-1} \mathbf{W}_{(t-1)r}(\beta) \mathbf{Z}'_{t-1} (y_{tr} - p_{tr}(\beta))$ and $\mathbf{W}_{(t-1)r}(\beta) = [\partial^2(l \circ h)/\partial \boldsymbol{\eta}_{t-1} \partial \boldsymbol{\eta}'_{t-1}]$.

4.2 Large Sample Theory

We still adhere to (A), the only exception being that the true model obeys (4.1) and the dimension pq now becomes $q + p$. Moreover, we assume that the Jacobian of $\mathbf{U}_{t-1}(\boldsymbol{\beta})$ is different than zero. We denote this modified version of assumption (A) by (B). The discussion after assumption (A) applies here too. In particular from (B.4) we see once again that the conditional information matrix has a limit in probability, that is

$$\frac{\mathbf{G}_N(\boldsymbol{\beta})}{N} \xrightarrow{p} \int_{R^{(q+p) \times p}} \mathbf{Z}\mathbf{U}(\boldsymbol{\beta})\boldsymbol{\Sigma}(\boldsymbol{\beta})\mathbf{U}'(\boldsymbol{\beta})\mathbf{Z}'d\nu(\mathbf{Z}) \equiv \boldsymbol{\Lambda}_2(\boldsymbol{\beta})$$

with $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ as before and $\mathbf{U}(\boldsymbol{\beta}) = [\partial(l \circ h)/\partial\boldsymbol{\eta}]$. As before $\boldsymbol{\Lambda}_2(\boldsymbol{\beta}) > \mathbf{0}$. We want to emphasize again that no Markovian assumption is necessary.

Theorem 4.1 Consider the proportional odds model and assume (B). Then:

1. The probability that a unique MPLE exists tends to 1 as $N \rightarrow \infty$.
2. The estimator $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &\xrightarrow{p} \boldsymbol{\beta}_0 \\ \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\xrightarrow{D} \mathcal{N}(0, \boldsymbol{\Lambda}_2^{-1}(\boldsymbol{\beta}_0)) \end{aligned}$$

Proof: Also here we only give a sketch of the proof. One can again prove under assumption (B) that the multivariate Central Limit Theorem for martingales applies to the partial score function,

$$\frac{S_N}{\sqrt{N}} \xrightarrow{D} \mathcal{N}(0, \boldsymbol{\Lambda}_2)$$

The next step, is to show that the negative matrix of the second derivatives (4.3) is “close”, in some sense, to the conditional information matrix, or equivalently

$$\frac{\mathbf{R}_N}{N} \xrightarrow{p} \mathbf{0} \tag{4.4}$$

This is verified using the boundedness of the covariates and the continuity of the conditional information matrix. Then we can prove again (3.8) by using Taylor expansion which establishes existence. Uniqueness can be proved along the lines of Burrige (1982). The main result there is that the integral of a log-concave function with respect to some of its

argument is a log-concave function of the remaining ones. Consistency follows in the same manner as before. Asymptotic normality can be proved by

$$S_N = \mathbf{H}_N(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx \mathbf{G}_N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

by appealing to the continuity of the conditional information matrix, the consistency of the MPLE, and equation (4.4). The conclusion of the theorem therefore holds. \square

5 Goodness of fit Statistic

A question which arises naturally after every procedure involving regression is that of goodness of fit. Our approach is to classify the responses \mathbf{y}_t according to mutually exclusive events in terms of the covariates \mathbf{Z}_{t-1} (see Schoenfeld (1980); Slud and Kedem (1994)). Since the theory we are going to develop applies to both models under consideration, we use a unified notation.

Suppose that A_1, \dots, A_k constitute a partition of $R^{p \times q}$, with p the appropriate dimension; i.e. either pq or $p + q$. For $l = 1, \dots, k$ define

$$M_l = \sum_{t=1}^N I_{[\mathbf{Z}_{t-1} \in A_l]} \mathbf{y}_t$$

and

$$E_l(\boldsymbol{\beta}) = \sum_{t=1}^N I_{[\mathbf{Z}_{t-1} \in A_l]} \mathbf{p}_t(\boldsymbol{\beta})$$

where I is the indicator of the set $\{\mathbf{Z}_{t-1} \in A_l\}$, for $l = 1, \dots, k$. Let $M_N = (M_1', \dots, M_k')'$, $E_N(\boldsymbol{\beta}) = (E_1'(\boldsymbol{\beta}), \dots, E_k'(\boldsymbol{\beta}))'$. If we let $I_{t-1} = (I_{[\mathbf{Z}_{t-1} \in A_1]}, \dots, I_{[\mathbf{Z}_{t-1} \in A_k]})'$ we can see that

$$\mathbf{d}_N(\boldsymbol{\beta}) = M_N - E_N(\boldsymbol{\beta}) = \sum_{t=1}^N I_{t-1} \otimes (\mathbf{y}_t - \mathbf{p}_t(\boldsymbol{\beta}))$$

with \otimes denotes the Kronecker product. It follows that $\mathbf{d}_N(\boldsymbol{\beta})$ is a zero mean square integrable martingale that satisfies all the conditions needed for an application of the Central Limit Theorem under our previous assumptions. It turns out that

$$\frac{\mathbf{d}_N}{\sqrt{N}} \xrightarrow{p} \mathcal{N}(0, \mathbf{C})$$

where $\mathbf{C} = \oplus_{l=1}^k \mathbf{C}_l$, the direct sum of k matrices¹, and \mathbf{C}_l is a $q \times q$ symmetric matrix given by

$$\mathbf{C}_l(\boldsymbol{\beta}_0) = \begin{bmatrix} \int_{A_l} p_1(\boldsymbol{\beta}_0)(1 - p_1(\boldsymbol{\beta}_0))d\nu(\mathbf{Z}) & \cdots & -\int_{A_l} p_1(\boldsymbol{\beta}_0)p_q(\boldsymbol{\beta}_0)d\nu(\mathbf{Z}) \\ & \ddots & \vdots \\ -\int_{A_l} p_1(\boldsymbol{\beta}_0)p_q(\boldsymbol{\beta}_0)d\nu(\mathbf{Z}) & \cdots & \int_{A_l} p_q(\boldsymbol{\beta}_0)(1 - p_q(\boldsymbol{\beta}_0))d\nu(\mathbf{Z}) \end{bmatrix}$$

From the above result we have the following proposition:

Proposition 5.1 *As $N \rightarrow \infty$, the asymptotic distribution of the statistic*

$$\chi^2(\boldsymbol{\beta}_0) = \frac{1}{N} \sum_{l=1}^k d'_l(\boldsymbol{\beta}_0) \mathbf{C}_l^{-1}(\boldsymbol{\beta}_0) d_l(\boldsymbol{\beta}_0) \quad (5.1)$$

is a chi-square with kq degrees of freedom.

Since $d'_l \mathbf{C}_l^{-1} d_l / N$ is distributed as chi-square with q degrees of freedom from the convergence of \mathbf{d}_N , we have that $\chi^2(\boldsymbol{\beta}_0)$ follows a chi-square with kq degrees of freedom as a sum of independent chi-square distributed random variables. This is so because the covariates belong to different partition sets, hence in the limit the components of \mathbf{d}_N are independent. The inverse of $\mathbf{C}_l(\boldsymbol{\beta}_0)$, $l = 1, \dots, k$ is guaranteed from either assumption (A.4) or (B.4).

6 An application

Due to various technological constraints, the effective dynamic range of a spaceborne precipitation radar (PR) flying at an altitude of 350 km is limited at present to intermediate values. In particular at high rain rates—the source of most of the rainfall volume—there is a degraded signal to noise ratio due to large attenuation (Meneghini and Jones (1993)). Basically this means the spaceborne PR saturates at some intermediate value (roughly 10-15 mm/hr) so that high rain rates are indistinguishable from lower rates. It is therefore useful to construct methods that can help a PR discern instantaneously high rain rates using covariate information.

In the following example it is shown that rain rate time series obtained by a PR can be classified instantaneously reasonably well using covariate information, and also that the instantaneous fraction of the area where it is raining (fractional area) is a useful covariate—in

¹ $\mathbf{A} \oplus \mathbf{B}$ creates a partitioned diagonal matrix, having \mathbf{A} , \mathbf{B} on the main diagonal.

line with Kedem and Pavlopoulos (1991). It stands to reason that additional useful covariates can improve the classification procedure. See Stern and Coe (1984) for a treatment of a similar problem.

Our data consist of two time series of length $N=643$ each. The first time series is the area average rain rate, the second the corresponding fractional area, both obtained from radar snapshots every 10 minutes (4.46 days) by means of a PR on board the ship R/V JV Vickers (USA). This is part of the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE) data set collected by shipborne Doppler radars during November 1992 - February 1993 in the China Sea (approximately 2^{deg}S and 156^{deg}E) over an area of roughly 300 km by 400 km. For a detailed description of the data see Short et al. (1995).

Let r_t denote the area average rain rate, and let x_t be the fractional area. We *categorize* rain rate in three and four bins. The new variable, say y_t , with three categories, is defined by the quantization

$$y_t = \begin{cases} 1 & \text{if } 0 \leq r_t < 0.005 \\ 2 & \text{if } 0.005 \leq r_t < 0.25 \\ 3 & \text{if } r_t \geq 0.25 \end{cases}$$

The variable with four categories, say \tilde{y}_t , is given by,

$$\tilde{y}_t = \begin{cases} 1 & \text{if } 0 \leq r_t < 0.005 \\ 2 & \text{if } 0.005 \leq r_t < 0.04 \\ 3 & \text{if } 0.04 \leq r_t < 0.25 \\ 4 & \text{if } r_t \geq 0.25 \end{cases}$$

These types of data are interval data; they can be taken as ordinal or nominal.

Table 6.1: Fitted Models

Model	Class	Covariates	Number of Categories
Model 1	MLM	$x_t, x_{t-1}, r_{t-3}, x_{t-4}r_{t-1}$	3
Model 2	POM	$x_t, x_{t-1}, r_{t-3}, x_{t-4}r_{t-1}$	3
Model 3	POM	$r_{t-1}, r_{t-2}, r_{t-3}, x_{t-2}, r_{t-1}x_t$	4
Model 4	POM	$r_{t-1}, r_{t-2}, r_{t-3}, x_t, x_{t-1}, r_{t-1}x_t$	4
Model 5	MLM	$x_{t-1}, r_{t-2}, r_{t-3}, x_tr_{t-1}$	4

Table 6.2: Probabilities of misclassification and goodness of fit test of the fitted models

Item	Model 1	Model 2	Model 3	Model 4	Model 5
First category	9.4%	7.5%	4.4%	5.6%	9.4%
Second Category	8.8%	12.5%	13.4%	9.1%	18.3%
Third Category	4.5%	3.2%	2.6%	1.9%	5.8%
Fourth Category	-	-	.6%	0%	1.9%
Overall	7%	8%	5.5%	4.2%	8%
χ^2	18.29	48.076	22.1713	10.482	20.703
d.f.	10	14	22	21	15
p-value	.05	.001	.44	.97	.146

The chi-square goodness of fit statistic is constructed from a partition with $k = 10$ cells obtained from the fractional area x_t . In calculating the p-values, the number of degrees of freedom is adjusted to account for the fact that (5.1) is computed with $\hat{\beta}$, the MPLE of β_0 . Table 6.1 gives the five models that were fitted. All the models consist of an intercept plus the indicated covariates in the third column of the table. The second column indicates whether the model is a multinomial logits model (MLM) or proportional odds model (POM). Model 1 and Model 2 were fitted to y_t . The rest of the models correspond to \tilde{y}_t . Table 6.2 gives the value of the chi-square test the corresponding p-values, degrees of freedom, and the probabilities of misclassification. For example of all the observations belonging to the first category of the second model, only 7.5% were misclassified etc. Evidently, Model 4 gives the best fit.

7 Summary

Two generalized logistic regression/autoregression models, that take into account random time dependent covariates, were presented. We saw that the partial likelihood inference is a promising method for estimating nonstationary one step transition probabilities. Martingale limit theory enabled us to prove large sample properties of the MPLE without reference to any Markov assumption. The use of a chi-square goodness of the fit test was illustrated using real data. An application reveals that the quality of fit depends on the number of categories, the boundaries and the set of covariates.

References

- Agresti, A. (1990) *Categorical Data Analysis*. New York: John Wiley.
- Andersen, K. P., Borgan, O., Gill, D. R. and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Billingsley, P. (1986) *Probability and Measure*. New York: John Wiley.
- Burridge, J. (1982) Some unimodality properties of likelihood derived from grouped data. *Biometrika* 69, 145–151.
- Cox, D. R. (1975) Partial likelihood. *Biometrika* 62, 69–76.
- Cox, D. R. and Snell, E. J. (1989) *The analysis of Binary Data* (2nd Edn). London: Chapman and Hall.
- Diggle, J. P., Liang, K.-Y. and Zeger, L. S. (1994) *Analysis of Longitudinal Data*. New York: Oxford University Press .
- Fahrmeir, L. and Kaufmann, H. (1987) Regression models for nonstationary categorical time series. *Journal of Time Series Analysis* 8, 147–160.
- Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.
- Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theorems and its Applications*. New York: Academic Press.
- Kaufmann, H. (1987) Regression models for nonstationary time series: Asymptotic estimation theory. *Annals of Statistics* 15, 79–98.
- Kedem, B. (1994) *Time Series Analysis by Higher Order Crossings*. New York: IEEE Press.
- Kedem, B. and Pavlopoulos, H. (1991) On the threshold method for rainfall estimation: Choosing the optimal threshold level. *Journal of the American Statistical Association* 86, 626–633.
- Korn, E. L. and Whittemore, A. S. (1979) Methods for analyzing panel studies of accute health effects of air pollution. *Biometrics* 35, 795–802.

- Mc Fadden, D. (1973) Conditional logit analysis of qualitative choice behavior. in P. Zarembka (ed.). *Frontiers in Econometrics*, New York: Academic Press.
- McCullagh, P. (1980) Regression models for ordinal data (with discussion). *Journal of Royal Statistical Society* B42, 109–142.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models* (2nd Edn). New York: Chapman and Hall.
- Meneghini, R. and Jones, A. J. (1993) An approach to estimate the areal rain-rate distribution from spaceborne radar by the use of multiple thresholds. *Journal of Applied Meteorology* 32, 386–398.
- Schoenfeld, D. (1980) Chi-square goodness-of-fit test for the proportional hazards regression model. *Biometrika* 67, 145–153.
- Short, D. A., Kucera, P. A., Ferrier, B. S. and Thiele, O. W. (1995) COARE IOP rainfall from shipborne radars: 1. Rain mapping algorithms. Submitted to the 27'th Conference on Radar Meteorology, Vail, Colorado.
- Slud, E. (1992) Partial likelihood for continuous time stochastic processes. *Scandinavian Journal of Statistics* 19, 97–109.
- Slud, E. and Kedem, B. (1994) Partial likelihood analysis of logistic regression and autoregression. *Statistica Sinica* 4, 89–106.
- Snell, E. J. (1964) A scaling procedure for ordered categorical data. *Biometrics* 20, 592–607.
- Stern, R. D. and Coe, R. (1984) A model fitting analysis of daily rainfall data. *Journal of Royal Statistical Society* A147, 1–34.
- Wong, W. H. (1986) Theory of partial likelihood. *Annals of Statistics* 14, 88–123.
- Zeger, S. L., Liang, K.-Y. and Self, S. G. (1985) The analysis of binary longitudinal data with time independent covariates. *Biometrika* 72, 8–31.