THE ANALYSIS OF LONGITUDINAL ORDINAL
RESPONSE DATA IN CONTINUOUS TIME

Michael R. Kosorok, PhD
Departments of Statistics and Biostatistics

Wei-Hsiung Chao
Department of Biostatistics

MADISON, WISCONSIN
The Analysis of Longitudinal Ordinal Response Data In Continuous Time

UNIVERSITY OF WISCONSIN-MADISON

DEPARTMENT OF BIOSTATISTICS TECHNICAL REPORT 87

Michael R. Kosorok and Wei-Hsiung Chao

ABSTRACT

A simple Markov model is developed for assessing the predictive effect of time-dependent covariates on an intermittently observed ordinal response in continuous-time. This is accomplished by reparameterizing an ergodic intensity matrix in terms of its equilibrium distribution and a parametrically independent component which assesses the rate of movement between ordinal categories. The effect of covariates on the equilibrium distribution can then be modeled using any link appropriate for ordinal data. A robust maximum likelihood estimator based on this model is shown to be consistent and asymptotically normal. Practical data analysis issues are discussed and a simple diagnostic tool for assessing model adequacy is developed. The utility of these methods is demonstrated with several analyses of visual acuity data, including a comparison analysis based on generalized estimating equation (GEE) methods.

KEY WORDS: Generalized estimating equations; Intensity matrix; Longitudinal data; Markov processes; Maximum likelihood estimation; Ordinal response; Regression analysis; Time-dependent covariates; Transitional model.

1Michael R. Kosorok is an Assistant Professor, Departments of Statistics and Biostatistics, University of Wisconsin, Madison, WI 53792. Wei-Hsiung Chao is a graduate student, Department of Statistics, University of Wisconsin, Madison, WI 53706. This research was supported in part by a University of Wisconsin Graduate School Award. The authors would like to thank Douglas M. Bates, Richard J. Chappell, David L. DeMets, Thomas G. Kurtz, and Michael A. Newton for helpful discussions on the methods proposed here. They also thank Marian R. Fisher and Stephen J. Gange for both providing and discussing the diabetic retinopathy study data.
1. INTRODUCTION

Many longitudinal studies involve taking repeated measurements of an ordinal response and several covariates from a sample of individuals. Examples of ordinal responses include visual acuity (ranging from better than 20/20 to nearly blind), Apgar score (ranging from 0 to 10, depending on the health of the infant), or even a binary indicator of the presence or absence of pain (a two-category response). The search for meaningful and statistically valid models for analyzing this kind of longitudinal data has been fairly intense in the last several decades. The depth and diversity of this research is evident in several excellent surveys on the topic, including the work of Agresti (1989), Ashby et al (1992), and Lindsey (1993).

Typically, there are two distinct approaches to modeling longitudinal data depending on whether scientific interest lies in the pattern of change in response over time or simply in the marginal dependence of a response on a set of covariates. The latter approach primarily focuses on marginal relationships between the covariates and the response while accounting for the correlation among the repeated observations for a given subject. Examples of this approach include the works of Wei and Johnson (1985), Liang and Zeger (1986), Stram, Wei, and Ware (1988), and Zeger, Liang, and Albert (1988). The former approach, however, focuses on predictive relationships between the current values of the response and past values of the covariates and response. In this setting, it is common to use Markov models for either discrete or continuous time. Examples of such models for discrete time include the logistic autoregressive regression model (Slud and Kedem, 1994) and the conditional log-linear model (Gilula and Haberman, 1994).

Several investigators have developed Markov models for continuous time. Kalbfleisch and Lawless (1985) described a general approach for applying Markov models to continuous time categorical data with time-independent covariates. This approach involves
modeling the intensity matrix which generates the categorical response as a function of covariates with, for example, each intensity element as a log-linear model in the covariates. Anderson, Hansen, and Keiding (1991) proposed a non-homogeneous Markov process for censored categorical longitudinal data in which proportional intensity regression is used to model each intensity element. The implicit assumption in this proportional intensity model is that the response is completely observed throughout the entire period of the study. Unfortunately, ordinal responses are frequently only assessed on an intermittent basis.

A common feature of both of these continuous time models is that each element of the intensity matrices is modeled as a separate function of the covariates. This can sometimes result in models which are complex and considerably more difficult to interpret than marginal models (Stram, Wei, and Ware, 1988). This is particularly true in settings where the actual categories are somewhat arbitrary, as is the case with ordinal data, where transitions between specific categories are frequently not as interesting as the general direction and rate of movement.

In this paper, we present a continuous time Markov model for intermittently observed ordinal data which permits unified inference on the relationship between time-dependent covariates and the general direction and rate of movement between ordinal categories. This is accomplished by decomposing an ergodic intensity matrix into its equilibrium distribution and a parametrically independent component which assesses the transfer rate between categories. The formulation of this model—the "Local Equilibrium Distribution" (LED) model—is given in Section 2. In Section 3, we develop a robust maximum likelihood estimation approach for this model which produces consistent and asymptotically normal parameter estimates. In Section 4, we discuss a simple diagnostic tool for assessing model appropriateness. The utility of these methods is evaluated with several analyses of visual acuity data from patients with diabetic
retinopathy—including a comparison analysis based on generalized estimating equation (GEE) methods—in Section 5. Finally, several summary and concluding remarks are given in section 6.

2. A LOCAL EQUILIBRIUM DISTRIBUTION (LED) MODEL

2.1 The Model Framework

Suppose individuals move independently among $K$ states, denoted by $k = 1 \ldots K$. Let $\mathcal{Y}(\cdot)$ be a $K$-dimensional ordinal response process in continuous time—for a randomly chosen individual—intermittently observed at a sequence of time points, with a 1 in the appropriate ordinal category and zeros elsewhere. Suppose we also observe a right continuous covariate process $\mathcal{Z}(\cdot)$ at possibly different time points from the observation of $\mathcal{Y}(\cdot)$. We assume that $\mathcal{Z}(\cdot)$ is constant between the times we observe it. We will assume that the response process $\mathcal{Y}(\cdot)$ is a first order Markov process generated by a piecewise constant intensity matrix which is a function of the process $\mathcal{Z}(\cdot)$ and several parameters. The $K \times K$ intensity matrix at time $t$, $Q(\theta, t)$, has entries $q_{ab}$ of the following form:

$$q_{ab} = \frac{\pi_b}{\pi_a + \pi_b} \mu_{(a \wedge b) (a \vee b)}, \quad a \neq b,$$

$$q_{aa} = - \sum_{b \neq a} q_{ab}, \quad a = 1 \ldots K,$$

for some strictly positive $\pi = \{\pi_1 \ldots \pi_K\}'$ such that $\pi_1 + \cdots + \pi_K = 1$ and some non-negative $\mu = \{\mu_{ab}, 1 \leq a < b \leq K\}'$ such that enough of the $\mu_{ab}$ values are strictly positive to ensure that $Q(\theta, t)$ is ergodic, and where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. The quantities $\pi$ and $\mu$ are modeled as distinct functions of $z = \mathcal{Z}(t)$ and parameter vector $\theta$.

It is not difficult to show that $\pi_a$ is the equilibrium probability for category $a$, $a = 1 \ldots K$, since the equilibrium distribution for an ergodic first order Markov process,
with intensity matrix $Q$, is the unique solution of $\pi'Q = 0$ such that $\pi_1 + \cdots + \pi_K = 1$ (Kalbfleisch and Lawless, 1985). The remaining quantities, $\mu_{ab}, 1 \leq a < b \leq K$, are non-negative real values which reflect the rates of transfer between pairs of categories. It is not difficult to show that this parameterization fully specifies all time reversible ergodic intensity matrices. However, because $Q(\theta, \cdot)$ is allowed to change whenever the covariate process $Z(\cdot)$ changes, both the time reversible property and the equilibrium distribution reparameterization are local phenomena: the actual resulting process $Y(\cdot)$ is not necessarily time reversible and is not in general stationary.

For these reasons, we will refer to this model as the local equilibrium distribution (LED) model for longitudinal ordinal data in continuous time. A similar kind of reparameterization is discussed by Tuma, Hannan, and Groeneveld (1979) for stationary birth-death processes and by Kosorok (1993) for non-stationary birth-death processes which generate longitudinal ordinal data. As we will describe later, when $Q(\theta, t)$—for a given time $t$—is reparameterized into $\pi$ and $\mu$, the $\pi$ term assesses the direction of movement (trend) while the $\mu$ terms assess the rate of movement (diffusion)—which is related to the autocovariance—in what is essentially a non-stationary AR(1) process.

There exists a number of intuitive models for assessing the effect of covariates on a marginal ordinal response distribution (see Agresti, 1989, for some examples). We can apply any one of these marginal models to describe the relationship between the covariate vector $z$ and the local equilibrium distribution $\pi$ at time $t$. One reasonable way of modeling the transfer rate vector $\mu$ is to assume that

$$\mu_{ab} = \exp[\xi_{ab} + \gamma'z],$$

$1 \leq a < b \leq K$: thus transfer rates are permitted to vary from pair to pair yet the overall influence of the covariates on rate of movement can be assessed by the single parameter vector $\gamma$.

The parsimony that can be achieved by this reparameterization, as an alternative to
modeling each intensity element as a separate function of covariates, can be substantial. For example, the four category setting has 12 intensity elements \( q_{ab} \), \( a = 1 \ldots 4 \) and \( b = 1 \ldots 4 \) such that \( a \neq b \), which could be modeled individually as

\[
q_{ab} = \exp[\alpha_{ab} + \beta_{ab}'z].
\]

This requires fitting and interpreting 12 different covariate coefficient vectors and estimating 12 constant terms. In contrast, the LED model with a cumulative logistic link for the local equilibrium distribution only requires fitting two covariate coefficient vectors (one for the trend and one for the diffusion) as well as estimating 3 cutpoint parameters (for the cumulative logistic link) and 6 additional constant terms (for the 6 transfer rates). If there are 5 covariates, this is a reduction from 72 parameters \((12 \times 5 + 12)\) to 19 parameters \((2 \times 5 + 3 + 6)\).

Before completing the formulation of the LED model, we need to introduce some additional notation. Let \( Y_i(\cdot) \) be the ordinal response process and \( Z_i(\cdot) \) the corresponding covariate process for individual \( i, i = 1 \ldots N \). Also for individual \( i \), define

\[
t_{ij0} \equiv \text{ordinal response observation times, } j = 0 \ldots n_i,
\]

\[
y_{ij} \equiv Y_i(t_{ij0}),
\]

\[
t_{ijl} \equiv \text{covariate and response observation times, } l = 0 \ldots m_{ij}, j = 0 \ldots n_i,
\]

and \( m_{ij} \geq 1 \) for \( j < n_i \);

\[
z_{ijl} \equiv Z_i(t_{ijl}), \text{ and}
\]

\[
Q_{ijl}(\theta) \equiv \text{intensity matrix at time } t_{ijl}, \text{ with local equilibrium distribution}
\]

\[
\pi^{ijl} \text{ and transfer rate vector } \mu^{ijl}, \text{ both of which are functions of } z_{ijl} \text{ and}
\]

\( \theta \) according to the LED model;

set \( t_{ij(m_{ij})} \equiv t_{i(j+1)0} \) for \( j = 0 \ldots n_i \) and \( m_{i(n_i)} \equiv 0 \); and assume \( n_i \geq 1 \). There is some redundancy here, in that \( z_{ij(m_{ij})} \equiv z_{i(j+1)0} \), but this will simplify some of the notation used later.
Let $X_i \equiv \{y_{i0}, t_{ijl}, x_{ijl}, l = 0 \ldots m_{ij}, j = 0 \ldots n_i\}$ and $Y_i = \{y_{ij}, j = 1 \ldots n_i\}$. Thus, if we condition on the initial ordinal response $y_{i0}$, $X_i$ is the observed design and covariate process while $Y_i$ is the observed ordinal response process for individual $i$, $i \ldots N$. If we assume that $\mathcal{Y}_i(\cdot)$ is generated according to the LED model, then

$$E[y_{ij+1}|y_{ij}, X_i] = \left(\prod_{l=1}^{m_{ij}} \exp[Q_{ij(l-1)}(\theta)\{t_{ijl} - t_{ij(l-1)}\}]\right)'y_{ij} \quad (1)$$

completely specifies the distribution of $Y_i$ conditional on $X_i$ by the properties of first order Markov processes and multinomial distributions.

2.2 The LED Model as a Non-stationary AR(1) Process

As mentioned earlier, the LED model is essentially a non-stationary continuous time autoregressive process of order 1. This follows from the fact that, for any ergodic and aperiodic intensity matrix $Q$ with equilibrium distribution $\pi$, any non-negative $K$-vector $y$ which sums to 1, and any $d \geq 0$,

$$y' \exp[Qd] = \exp[-\lambda d]y'B(d) + (1 - \exp[-\lambda d])\pi', \quad (2)$$

where $\lambda > 0$ and $B(d)$ is a bounded matrix with all rows summing to 1 and such that $B(0) = I$. The proof of this result is given in the appendix.

If, for example, $Q$ is parameterized according to the LED model and the elements of $\mu$ are restricted to be $\mu_{ab} = (\pi_a + \pi_b)\mu$, for $1 \leq a < b \leq K$, then the decomposition given by (2) becomes simply

$$y' \exp[Qd] = \exp[-\mu d]y' + (1 - \exp[-\mu d])\pi', \quad (2)$$

which clearly has a non-stationary AR(1) structure.

In general, the result given in (2) implies that the conditional expectation (1) can be reexpressed as an exponentially weighted average of local equilibrium distributions and $y_{ij}$. To see this, first let $\lambda_{ijl}, B_{ijl}(\cdot)$, and $\pi_{ijl}$ be the terms corresponding to the
decomposition given in (2) for the intensity matrix \(Q_{ij}(\theta)\), and define \(d_{ijl} \equiv t_{ij(l+1)} - t_{ijl}\). Now, the conditional expectation (1) can be reexpressed as

\[
\begin{align*}
E[y_{i(j+1)}|y_{ij}, X_i] &= \exp \left[ -\sum_{l=1}^{m_{ij}} \lambda_{ij(t-1)}d_{ij(t-1)} \right] \left( \prod_{l=1}^{m_{ij}} B_{ij(t-1)}(d_{ij(t-1)}) \right)' y_{ij} \\
&+ \left( 1 - \exp \left[ -\lambda_{ij(m_{ij}-1)}d_{ij(m_{ij}-1)} \right] \right) \pi^{ij(m_{ij}-1)} \\
&+ I_{\{m_{ij} > 1\}} \sum_{l=1}^{m_{ij}-1} \left( 1 - \exp \left[ -\lambda_{ij(l-1)}d_{ij(l-1)} \right] \right) \\
&\times \exp \left[ -\sum_{r=i+1}^{m_{ij}} \lambda_{ij(r-1)}d_{ij(r-1)} \right] \left( \prod_{r=i+1}^{m_{ij}} B_{ij(r-1)}(d_{ij(r-1)}) \right)' \pi^{ij(l-1)},
\end{align*}
\]

where

\[
I_{\{C\}} = \begin{cases} 
1, & \text{if } C \text{ is true} \\
0, & \text{otherwise.}
\end{cases}
\]

Thus, the LED model can also be interpreted as a fairly rich class of continuous time non-stationary AR(1) models, with the trend terms being the local equilibrium distributions and the autocovariance terms reflecting the transfer rates.

### 2.3 A Special Class of LED Models

For some ordinal outcomes, it may be appropriate to assume that instantaneous transitions are restricted to adjacent or nearly adjacent categories. This would be the case, for example, if progression from low to high categories was thought to be gradual. This kind of restriction is equivalent to requiring that the elements of the transfer rate vector \(\mu\)—from the LED parameterization of the intensity matrix \(Q\)—be zero for transitions between states more than some \(d \geq 1\) categories apart. Such a restriction results in an intensity matrix which is \((2d + 1) \land K\)-diagonal. For a given value of \(d\), we will refer to this restricted model as a “\(d\)-level local equilibrium distribution” (LED-\(d\)) model.

For example, an LED-1 model is a simple birth-death process with boundaries 1 and \(K\). This model would be appropriate when individuals are thought to move a
distance of only one category at a time. It turns out that the class of LED-1 models
fully specifies all ergodic, aperiodic Markov processes with boundaries 1 and K defined
by a tridiagonal intensity matrix. This follows from the fact that all ergodic birth-death
processes are time reversible (Ross, 1993).

The binary, or two category, case is the simplest of all LED-1 models. In fact, the
corresponding 2-vector \( y \) can be replaced with a dichotomous variable \( y^* \), where

\[
 y^* = \begin{cases} 
 0, & \text{for } y = \{1, 0\}', \\
 1, & \text{for } y = \{0, 1\}'. 
\end{cases}
\]

With this substitution, the conditional expectation (3) becomes

\[
 E[y^*_{i(j+1)}|y^*_i, X_i] = \exp \left[ - \sum_{l=1}^{m_{ij}} \mu^{ij(l-1)} d_{ij(l-1)} \right] y^*_{ij} + \left( 1 - \exp \left[ - \mu^{ij(m_{ij}-1)} d_{ij(m_{ij}-1)} \right] \right) \pi^*_{ij(m_{ij}-1)} \\
 + I_{\{m_{ij}>1\}} \sum_{l=1}^{m_{ij}} \left( 1 - \exp \left[ - \mu^{ij(l-1)} d_{ij(l-1)} \right] \right) \exp \left[ - \sum_{r=l+1}^{m_{ij}} \mu^{ij(r-1)} d_{ij(r-1)} \right] \pi^*_{ij(l-1)},
\]

where \( \pi^*_{ij} \) is the local equilibrium probability of being in category 2 which corresponds
to the \( 2 \times 2 \) intensity matrix \( Q_{ijl}(\theta) \), and where \( \mu^{ijl} \) is the associated transfer rate
between states 1 and 2.

3. ESTIMATION AND LARGE SAMPLE THEORY

3.1 The Estimation Procedure

Before outlining the maximum likelihood estimation procedure and corresponding
large sample characteristics, we need to define several quantities. Let

\[
 \ell_{ij}(\theta) = \log \left\{ y^*_i \left( \prod_{l=1}^{m_{ij}} \exp[Q_{ij(l-1)}(\theta)\{t_{ijl} - t_{ij(l-1)}\}] \right) y^*_{i(j+1)} \right\}
\]

be the \( i \)'th individual's contribution to the likelihood on the interval \( (t_{ij0}, t_{ij(j+1)0}] \) de-

fined by the conditional expectation (1); and let the associated contribution to the
score vector be \( s_{ij}(\theta) \) with elements

\[
 s_{ij}^u(\theta) = \frac{\partial \ell_{ij}(\theta)}{\partial \theta_u},
\]
$u = 1 \ldots M$, and where $\theta_u$ is the $u$'th element of the parameter $M$-vector $\theta$.

Our procedure for estimating the model parameter vector $\theta$ is to solve the score equation $S_N(\hat{\theta}_N) = N^{-1} \sum_{i=1}^{N} \sum_{j=0}^{n_i-1} s_{ij}(\hat{\theta}_N) = 0$. As shown by Kalbfleisch and Lawless (1985), the required first derivatives can be obtained directly from the canonical decomposition and first derivatives of $Q_{ij}(\theta)$. In fact, as will be shown in section 3.2, second and higher derivatives can also obtained from the canonical decomposition and derivatives of $Q_{ij}(\theta)$. The existence and behavior of these derivatives will be utilized in section 3.3 to obtain asymptotic consistency and normality for $\hat{\theta}_N$ under fairly reasonable regularity conditions, even when the LED based likelihood is not correct for the data at hand but is reasonably close.

This estimation procedure begins with determining an initial estimate of $\theta$ by assuming that all covariate effects are zero; that the cut-point parameters for the local equilibrium distributions are such that $\pi^{ij} = \{1/K, 1/K, \cdots, 1/K\}'$ for all $i$, $j$, and $l$; and that $\mu^{ij} = \{\mu, \mu, \cdots, \mu\}'$ for all $i$, $j$, and $l$, where $\mu > 0$ is very small. Under these assumptions, an approximate estimate of $\mu$ is simply

$$
\hat{\mu} = \frac{\sum_{i=1}^{N} \sum_{j=0}^{n_i-1} x_{ij}}{(K - 1) \sum_{i=1}^{n_i-1} \delta_{ij}},
$$

where $\delta_{ij} = t_{i(j+1)0} - t_{ij0}$ and $x_{ij} = I(y_{ij} \neq y_{i(j+1)})$. This estimate is simply an average transition rate per unit time and is approximately unbiased for $\mu$, provided $\mu$ is small enough so that at most one change can occur in the interval $(t_{ij0}, t_{i(j+1)0}]$ for all $i$ and $j$.

Beginning with this initial estimate $\hat{\theta}_N^{(0)}$, a Fisher-scoring algorithm with updates of the form

$$
\hat{\theta}_N^{(n+1)} = \hat{\theta}_N^{(n)} + \left( \sum_{i=1}^{N} \sum_{j=0}^{n_i-1} s_{ij}(\hat{\theta}_N^{(n)}) \{s_{ij}(\hat{\theta}_N^{(n)})\}' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{j=0}^{n_i-1} s_{ij}(\hat{\theta}_N^{(n)}) \right)
$$

is used several times to get into the neighborhood of $\hat{\theta}_N$, after which the exact second derivative is employed in several Newton-Raphson iterations until convergence.
3.2 The Computation of Derivatives

The greatest difficulty in computing derivatives of the log-likelihood components \( l_{ij} \), \( j = 0 \ldots n_i, \ i = 1 \ldots N \), is computing the derivatives of \( \exp[Q_{ij(l-1)}(\theta)\{t_{ijl} - t_{ij(l-1)}\}] \), \( l = 1 \ldots m_{ij} \). Kalbfleisch and Lawless (1985) showed that for a general intensity matrix \( Q \) with canonical decomposition \( A\DA^{-1} \) and a time change \( t \),

\[
\frac{\partial \exp[Qt]}{\partial \theta_u} = \DA \hat{V}_u A^{-1},
\]

for \( u = 1 \ldots M \), where \( \hat{V}_u \) is a \( K \times K \) matrix with \((a, b)\)’th element

\[
(\hat{g}_{ab}^u) \frac{\exp[d_a t] - \exp[d_b t]}{d_a - d_b} \text{ for } d_a \neq d_b, \text{ or } (\hat{g}_{ab}^u t) \exp[d_a t] \text{ for } d_a = d_b,
\]

where \( d_a \) is the \( a \)'th diagonal element of \( D \) and where \( \hat{g}_{ab}^u \) is the \((a, b)\)’th element of \( A^{-1}(\partial Q/\partial \theta_u)A \).

As shown in the appendix, a similar result can be obtained for the second derivative:

\[
\frac{\partial^2 \exp[Qt]}{\partial \theta_u \partial \theta_v} = \DA \hat{V}_{uv} A^{-1} + \DA \hat{H}_{uv} A^{-1}, \tag{4}
\]

for \( u = 1 \ldots M \) and \( v = 1 \ldots M \), where \( \hat{V}_{uv} \) is a \( K \times K \) matrix with \((a, b)\)’th element

\[
(\hat{g}_{ab}^{uv}) \frac{\exp[d_a t] - \exp[d_b t]}{d_a - d_b} \text{ for } d_a \neq d_b, \text{ or } (\hat{g}_{ab}^{uv} t) \exp[d_a t] \text{ for } d_a = d_b,
\]

where \( \hat{g}_{ab}^{uv} \) is the \((a, b)\)’th element of

\[
A^{-1}\left(\frac{\partial^2 Q}{\partial \theta_u \partial \theta_v}\right)A;
\]

and where \( \hat{H}_{uv} \) is a \( K \times K \) matrix with \((a, b)\)’th element

\[
\sum_{r=1}^{K} [\hat{g}_{ar}^u \hat{g}_{rb}^v + \hat{g}_{ar}^v \hat{g}_{rb}^u] \Gamma_{arb},
\]
where
\[
\Gamma_{arb} = \begin{cases} 
\frac{e^{d_a t} - d_a t - 1}{(d_a - d_b)(d_a - d_r)} + \frac{e^{d_b t} - d_b t - 1}{(d_b - d_a)(d_b - d_r)}, & \text{if } d_a \neq d_b, \ d_a \neq d_r, \ d_b \neq d_r, \\
\frac{t(e^{d_a t - 1})}{d_a - d_r} + \frac{e^{d_r t - d_a t + d_a t}}{(d_a - d_r)^2}, & \text{if } d_a = d_b \neq d_r, \\
\frac{t(e^{d_a t - 1})}{d_a - d_a} + \frac{e^{d_r t - d_a t + d_a t}}{(d_a - d_a)^2}, & \text{if } d_a = d_r \neq d_b, \\
\frac{t(e^{d_b t - 1})}{d_b - d_b} + \frac{e^{d_a t - d_b t + d_b t}}{(d_b - d_b)^2}, & \text{if } d_b = d_r \neq d_a, \\
\frac{t^2}{2} e^{d_a t}, & \text{if } d_a = d_b = d_r.
\end{cases}
\]

Although complex, this result is easily computed; furthermore, close inspection of the terms of these derivatives reveals that they are bounded whenever the eigenvalues, $t$, and the derivatives of $Q$ are bounded. It can be shown that higher derivatives have similar structures and are similarly bounded.

An essential part of obtaining these derivatives involves obtaining the canonical decomposition of $Q$. This canonical decomposition is also useful for computing model-based predictions. If we assume that $Q$ is parameterized according to the LED model with local equilibrium distribution $\pi$, then certain simplifications in the canonical decomposition are possible. Let $c = \{\sqrt{\pi_1}, \sqrt{\pi_2}, \ldots, \sqrt{\pi_K}\}'$ and let $C$ be a $K \times (K - 1)$ matrix with $(a, b)'$th element $w_a/w_{a-1}$ if $a = b$, $-\sqrt{\pi_a \pi_b}/(w_{a}w_{b-1})$ if $a > b$, and 0 if $a < b$, where $w_0 = 1$ and, for $a \geq 1$,
\[
w_a = \sqrt{1 - \sum_{r=1}^{a} \pi_r}.
\]

If we let $D_c = \text{diag}(c)$, it is not difficult to see that $D_c Q D_c^{-1}$ is symmetric and therefore has real eigenvalues and orthonormal eigenvectors. Further, since $Q$ is time reversible, we know that one eigenvalue is 0 and the remaining eigenvalues are strictly negative.

By computing several inner products, it is evident that $c$ is the eigenvector of $D_c Q D_c^{-1}$ corresponding to the 0 eigenvalue and that the remaining eigenvalues are orthonormal linear combinations of the columns of $C$, since $C'c = 0$ and $C'C$ is the
identity matrix. Thus \( M = C'D_cQD_c^{-1}C \) is a symmetric matrix of order \( K - 1 \) with eigenvalues equal to the nonzero eigenvalues of \( Q \). Furthermore, it can be shown that the structure of \( C \) ensures that if \( Q \) has bandwidth \( m \), then so does \( M \). This means that computationally efficient algorithms for symmetric banded matrices can be fully utilized (see Golub and Van Loan, 1989). Now, if we let \( M = RAR' \) be the canonical decomposition of \( M \), then the full canonical decomposition of \( Q \) can be written \( Q = WFW^{-1} \), where

\[
W = D_c^{-1}[c:CR], \quad F = \begin{bmatrix} 0 & 0 \\ \vdots & \ddots \\ 0 & \Lambda \end{bmatrix}, \quad \text{and} \quad W^{-1} = \begin{bmatrix} c' \\ \vdots & \ddots \\ R'C' \end{bmatrix} D_c.
\]

This device allows us to take advantage of our knowledge of the zero eigenvalue in a computationally efficient manner.

3.3 Large Sample Results

In this section, we present robust large sample results which take advantage of the existence of the derivatives described in the previous section. We will begin with a theorem and corollary giving the main results followed by a brief discussion of how these regularity conditions should be satisfied in practical data analysis settings and under what circumstances the resulting model estimates are robust:

**Theorem 1** Let \( \Theta \) be an open convex set and suppose the following conditions hold:

1. The random pairs \( \{X_i, Y_i\}, i = 1 \ldots N, \) are independent and identically distributed.

2. For every \( \theta \in \Theta \), \( \ell_i(\theta) = \sum_{j=0}^{n_i-1} \ell_{ij}(\theta) \) is three times differentiable with respect to \( \theta \).

3. For every compact set \( T \subset \Theta \), there exists a function \( g(\cdot, \cdot) \) such that \( E[g(X_i, Y_i)] < \infty \) and, for every \( \theta \in T \),
(a) \( |\ell_i(\theta)| \leq g(X_i, Y_i) \),

(b) \( \frac{\partial s_i(\theta)}{\partial \theta_c} \leq g(X_i, Y_i) \),

(c) \( \left| \frac{\partial^2 s_i(\theta)}{\partial \theta_c \partial \theta_v} \right| \leq g(X_i, Y_i) \),

(d) \( |s_i^u(\theta)s_i^v(\theta)| \leq g(X_i, Y_i) \), and

(e) \( \left| \frac{\partial s_i^u(\theta)}{\partial \theta_v} s_i^v(\theta) \right| \leq g(X_i, Y_i) \),

where \( s_i^u(\theta) = \sum_{j=0}^{n_i-1} s_j^u(\theta) \), for \( u = 1 \ldots M \), \( v = 1 \ldots M \), and \( w = 1 \ldots M \).

4. \( S(\theta) = E[s_i(\theta)] \) has a unique zero in \( \Theta \) at \( \theta = \theta_0 \).

5. There exists an \( N^* < \infty \) such that for every \( N \geq N^* \),

\[
S_N(\theta) = N^{-1} \sum_{i=1}^{N} s_i(\theta)
\]

has a unique zero in \( \Theta \) at \( \theta = \hat{\theta}_N \).

6. For every \( \epsilon > 0 \), there exists a convex compact set \( T^* \subset \Theta \) and an \( N^{**} < \infty \) such that \( \theta_0 \in T^* \) and, for every \( N \geq N^{**} \), \( \Pr\{\hat{\theta}_N \in T^*\} > 1 - \epsilon \).

7. There exists a neighborhood of \( \theta_0 \), \( B(\theta_0) \), such that the matrix \( I(\theta) \), with \( (u,v)' \)th element

\[
I_{uv}(\theta) = E\left[ \frac{\partial s_i^u(\theta)}{\partial \theta_v} \right],
\]

is non-singular for all \( \theta \in B(\theta_0) \).

Then the following are true:

1. \( \hat{\theta}_N \to \theta_0 \) in probability, as \( N \to \infty \).

2. \( \left[ I_N(\hat{\theta}_N) \right]^{-1} V_N(\hat{\theta}_N) \left[ I_N(\hat{\theta}_N) \right]^{-1} \to \left[ I(\theta_0) \right]^{-1} V(\theta_0) \left[ I(\theta_0) \right]^{-1} \) in probability, as \( N \to \infty \), where \( I_N(\theta) \), \( V_N(\theta) \), and \( V(\theta) \) are matrices with \( (u,v)' \)th elements

\[
I_{uv}^N(\theta) = -N^{-1} \sum_{i=1}^{N} \frac{\partial s_i^u(\theta)}{\partial \theta_v},
\]
\[ V_N^{uv}(\theta) = N^{-1} \sum_{i=1}^{N} s_i^u(\theta) s_i^u(\theta), \text{ and} \]
\[ V^{uv}(\theta) = \mathbb{E}[s_i^u(\theta) s_i^u(\theta)], \]
respectively, for \( u = 1 \ldots M \) and \( v = 1 \ldots M \).

3. \( N^{1/2}(\hat{\theta}_N - \theta_0) \to \mathbf{Z} \) in distribution, as \( N \to \infty \), where

\[ \mathbf{Z} \sim N \left( \mathbf{0}, [\mathbf{I}(\theta_0)]^{-1} \mathbf{V}(\theta_0) [\mathbf{I}(\theta_0)]^{-1} \right). \]

**Corollary 1** If, in addition to the conditions specified in theorem 1, we have that \( \ell_i(\theta_0) \) is the true log conditional likelihood of \( \mathbf{Y}_i \) given \( \mathbf{X}_i \), for \( i = 1 \ldots N \), then the following are true:

1. \( \hat{\theta}_N \to \theta_0 \) in probability as \( N \to \infty \).

2. \( [\mathbf{I}_N(\hat{\theta}_N)]^{-1} \to [\mathbf{I}(\theta_0)]^{-1} \) in probability as \( N \to \infty \).

3. \( \mathbf{V}_N(\hat{\theta}_N) \to \mathbf{V}(\theta_0) = \mathbf{I}(\theta_0) \) in probability as \( N \to \infty \).

4. \( N^{1/2}(\hat{\theta}_N - \theta_0) \to \mathbf{Z}^* \) in distribution, as \( N \to \infty \), where \( \mathbf{Z}^* \sim N \left( \mathbf{0}, [\mathbf{I}(\theta_0)]^{-1} \right). \)

Condition 6 serves a purpose similar to Lemma 2 of Huber (1967), but Huber’s approach otherwise differs substantially from ours.

Under a number of practical scenarios, the conditions of theorem 1 should be satisfied. For example, if we are willing to assume that each \( \pi^{ijl} \) component of the intensity matrices \( \mathbf{Q}_{ijl} \) is the cumulative logit of \( \beta' \mathbf{z}_{ijl} \) (with the necessary cutpoint parameters); each \((a,b)\)’th element of \( \mu^{ijl}, \mu_{ab}^{ijl} \), is \( \exp[\xi_{ab} + \gamma' \mathbf{z}_{ijl}] \); the moment generating function of the time-dependent covariates \( \mathbf{z}_{ijl} \) exists; and \( \sum_{j=0}^{n-1} m_{ij} \) is bounded, then for any compact \( T \in \Theta, \mathbf{Q}_{ijl} \) and the first three derivatives of \( \mathbf{Q}_{ijl} \) are bounded in expectation on \( T \). From the results of section 3.2, this means that \( \exp[\mathbf{Q}_{ijl}t] \) and its first three
derivatives are also bounded, satisfying conditions 2 and 3. If in addition we are willing to assume that individuals are independent and identically distributed and that \( S_N(\theta) \) is sufficiently well behaved so that

- its expectation has a unique zero in the interior of \( \Theta \) at some finite \( \theta_0 \),

- \( \hat{\theta}_N \) is unique in the interior of \( \Theta \) and does not drift off to infinity as \( N \to \infty \), and

- Its derivative is full rank when \( \theta \) is sufficiently close to \( \theta_0 \),

then conditions 1, 4, 5, 6, and 7 are also satisfied and the results of theorem 1 will follow.

In addition, theorem 1 does not require that the working LED likelihood be correct for the data at hand. However, the actual likelihood for the data should be close enough to the working likelihood that the resulting estimators are consistent for meaningful parameters. Whether or not the equilibrium model is “close enough” for a given setting is a complex question. As pointed out earlier, the LED model generates a non-stationary AR(1) process. If the structure of the trend and diffusion terms for the resulting non-stationary AR(1) process seem appropriate for a given analysis setting, then the estimation procedure given here should provide estimates of meaningful parameters whether or not the LED model is precisely correct.

4. A SIMPLE DIAGNOSTIC FOR MODEL APPROPRIATENESS

A simple diagnostic can be generated from the conditional expectation based on the LED model. First let \( p_{ij}(\theta) \) be the right hand side of (1), then if the LED model is correct, \( E[Y_{ij}|Y_{i(j-1)}, X_i] = p_{ij}(\theta_0) \), and \( \text{var}[Y_{ij}|Y_{i(j-1)}, X_i] = \text{diag}\{p_{ij}(\theta_0)\} - p_{ij}(\theta_0) \{p_{ij}(\theta_0)\}' \). If we now define \( v = \{1, 2, \ldots, K\}' \), then \( v'y_{ij} \) is the ordinal category which individuals \( i \) is in at time \( t_{ij0} \). Furthermore, the quantity

\[
r_{ij}(\theta) = \frac{v'y_{ij} - p_{ij}(\theta)}{\sigma_{ij}(\theta)},
\]

16
where

\[ \sigma_{ij}^2(\theta) = \mathbf{v}' [\text{diag} \{ \mathbf{p}_{ij}(\theta_0) \}] \mathbf{v} - \{ \mathbf{v}' \mathbf{p}_{ij}(\theta) \}^2, \]

has mean 0, variance 1, and is uncorrelated with each \( r_m(\theta), m \neq j \), when the LED model is correct and \( \theta = \theta_0 \). The same properties should be approximately true when \( \theta_0 \) is replaced with \( \hat{\theta}_N \). We will hereafter refer to \( r_{ij}(\hat{\theta}_N) \) as the “summary residual.” Scatterplots of this diagnostic against covariates can be used to assess a number of assumptions, including the correctness of the functional form of the covariates.

5. SEVERAL ANALYSES OF VISUAL ACUITY DATA

5.1 The Visual Acuity Data

The Diabetic Retinopathy Study (DRS) was a randomized controlled clinical trial initiated in 1972 to evaluate the effectiveness of photocoagulation treatment in preventing severe visual loss in patients with diabetic retinopathy. Eligible patients were randomized so that one eye received treatment and one did not. The patients were scheduled to return every 4 months to assess visual acuity and take several relevant measurements. On the basis of statistically significant evidence that photocoagulation therapy was effective in reducing the risk of severe visual loss in patients with advanced diabetic retinopathy, the protocol changed in 1976 to permit the treatment of initially untreated eyes which had “high risk of severe visual loss”. For this initial assessment, severe visual loss was defined as having visual acuity worse than 5/200 for at least two consecutively completed four-month follow-up visits. Follow-up continued into 1979 (Diabetic Retinopathy Study Research Group, 1976 & 1981; Rand et al, 1985).

In the analyses presented here, we will use data only from the right eye. Visits where either the covariates or response values were unavailable will be considered missing and the covariate value at the missing value will be assumed to be unchanged from the previous observed value (missing was actually extremely rare). This is not unreasonable
if we consider the time-dependent covariates to be as-observed values. Since covariates and responses were observed at the same time, \( m_{ij} = 1 \) for all \( i \) and \( j \). For this reason, the \( l \) subscript will not be needed and will be dropped from the analyses presented here. There are more than 1600 subjects included in this data set with an average of over 10 follow-up visits per subject. The analyses presented here were implemented in Fortran 77 on a Sun SPARCstation IPX.

The covariates for this analysis consisted of four baseline covariates and six time-dependent covariates in addition to the intercept term. The baseline covariates (abbreviations given in parentheses) were age at enrollment in years, rounded to the nearest integer (age); patient gender, 0 if male, 1 if female (sex); baseline retinal group, taking on the value 1 if at least some new blood vessels were present on the optic disc of the right eye at enrollment, 0 otherwise (BRG); and treatment assignment, 1 if the right eye initially received treatment, 0 if not (treatment). The six time-dependent covariates (abbreviations again given in parentheses) were time in the study since enrollment (time); protocol change status, taking on the value 1 for the period beginning with the visit in which the protocol changed, 0 otherwise (change); time by change interaction \((t \times c)\); treatment by time interaction \((T \times t)\); treatment by change interaction \((T \times c)\); and treatment by time by change interaction \((T \times t \times c)\).

5.2 A Binary LED-1 Analysis

For the binary analysis presented here, we used the response

\[
y_{ij}^* = \begin{cases} 
1 & \text{if visual acuity at time } t_{ij} \text{ is } > 5/200, \\
0 & \text{otherwise},
\end{cases}
\]

for \( j = 1 \ldots n_i, i = 1 \ldots N \), since this was the definition of severe visual loss as given in the original protocol. The logit link of \( \beta'z_{ij} \) was used for the local equilibrium probabilities \( \pi_{ij}^* \), while the transfer rates \( \mu^{ij} \) were assumed to be all equal, \( \mu^{ij} = \mu = \exp[\xi] \) (and therefore not dependent on covariates). We also need to define \( \phi = \exp[-\mu] \).