RT-MAE 9802

LONGITUDINAL DATA ESTIMATING EQUATIONS FOR DISPERSION MODELS

by

Rinaldo Artes
and
Bent Jørgensen

Palavras-Chave: Directional data, generalized estimating equations, generalized linear models, quasi-likelihood, simplex distribution, von Mises distribution.

Classificação AMS: 62F99, 62J99.
(AMS Classification)

- Março de 1998 -

Longitudinal Data Estimating Equations for Dispersion Models

Rinaldo Artes
Department of Statistics, University of São Paulo, Brazil
Bent Jørgensen
Department of Statistics, University of British Columbia, Canada

Abstract

We extend Liang and Zeger's generalized estimating equation approach for longitudinal data analysis to the case of marginal distributions from the class of dispersion models. Special cases include the von Mises and simplex distributions, suitable for angles and proportions, respectively. We consider modelling of position as well as joint modelling of position and dispersion.

Key words and phrases: directional data, generalized estimating equations, generalized linear models, quasi-likelihood, simplex distribution, von Mises distribution.

1 Introduction

The generalized estimating equation (GEE) method for analysis of longitudinal data was introduced by Liang and Zeger (1986) and Zeger and Liang (1986). This method is based on the quasi-likelihood approach of Wedderburn (1974), and provides estimating equations for the analysis of longitudinal regression models based on second-moment assumptions for the response variable.

The GEE method was motivated by ideas from natural exponential families, and was first applied mainly to discrete data, but many other types of applications followed. The paper Zeger, Liang and Albert (1988) considered random effects models, whereas Li (1994) and Zeger and Qaquish (1988) considered time series models. Prentice (1988) and Prentice and Zhao (1991) included the estimation of correlation (nuisance) parameters in the estimating equations. Applications of GEE may be found in Albert and McShane (1995) (analysis of spatially correlated data), Whittemore and Gong (1994) (case-control studies), Kenward et al. (1994) (ordinal longitudinal data), Liang and Hanfelt (1994) (teratological experiments), Cologne et al. (1993) (Poisson data) and Miller et al. (1993) (longitudinal polytomous data), among others. The book by Diggle et al. (1994) and the papers Liang, Zeger and Qaquish (1992) and Fitzmaurice et al. (1993) provide good overviews of the area.

In spite of this multitude of applications of the GEE method, certain types of data are not easily analysed by these methods. We propose an extension based on the class of dispersion models (Jørgensen, 1997a,b) that caters for many different types of non-normal distributions outside the class of natural exponential families, such as for example models for angles and proportions. The main idea is to replace the score function $(Y - \mu)/V(\mu)$ appearing in the generalized estimating equation by the score function $u(Y;\mu)$ from a dispersion model. The method is thus based on working second-moment assumptions for $u(Y;\mu)$ rather than for Y. The standard errors of the estimators may be consistently estimated by an empirical sandwich estimator.

Just as the original GEE method, the method is useful for estimation of regression parameters in situations where the marginal distributions of the data are known, but not the joint distribution. Few multivariate probability models with the required marginals exist, and even when they do, estimation may be hard, making GEE methods come in handy in many situations.

The set-up is introduced in Section 2, and in Section 3 we review some important results for estimating functions. Section 4 deals with position modelling, and Section 5 with position and dispersion modelling. A simulation study is presented in Section 6, and in Section 7 we consider an application to bird orientation data.

2 Dispersion models and longitudinal data

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ denote the response vector for the ith of n independent subjects, where $t = 1, \ldots, n_i$ indicates time. Let $\mathbf{x}_{it}$ denote a $p \times 1$ vector of covariates for Y_{it}. Assume a dispersion model DM for the marginal distributions of the observations,

$$Y_{it} \sim \mathrm{DM}(\mu_{it}; \sigma^2/w_{it}), \;\; \mu_{it} = h\left(\mathbf{x}_{it}^\top \boldsymbol{\beta}\right),$$

for all i and t, where w_{it} are known positive weights, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters and the inverse link function h is one-to-one and twice differentiable.

A univariate dispersion model $\mathrm{DM}(\mu, \sigma^2)$ (Jørgensen, 1997b) is defined by the probability density function

$$f_Y(y) = a(y;\sigma^2) \exp\left\{-\frac{1}{2\sigma^2} d(y;\mu)\right\}, \;\; \mu \in \Omega, \;\; \sigma^2 > 0, \tag{1}$$

where Ω is an interval and d is a unit deviance, that is, $d(y;\mu)$ is nonnegative, and zero if and only if $y = \mu$. The parameter μ is called the *position parameter* and σ^2 the *dispersion parameter*. We assume that the model is regular with respect to inference on the parameters μ and σ^2.

In the following we often use the index or concentration parameter $\lambda = 1/\sigma^2$ instead of σ^2, when convenient. In this case the function $a(y;\sigma^2)$ is written as $\exp\{c(y;\lambda)\}$.

An important special case is the class of *exponential dispersion models* where the unit deviance takes the form

$$d(y;\mu) = y g_1(\mu) + g_2(\mu) + g_3(y),$$

for suitable functions g_1, g_2 and g_3. When σ^2 is known, this gives a natural exponential family, including for example the binomial and Poisson families. The gamma, normal and inverse Gaussian families are examples of exponential dispersion models. These three families are members of the class of Tweedie exponential dispersion models, mainly suitable for positive data, corresponding to variance functions of the form μ^p for some $p \in (-\infty, 0] \cup [1, \infty)$, see Jørgensen (1997b, Ch. 4).

The class of *proper dispersion models* is defined by $a(y;\sigma^2)$ factorizing as $a(\sigma^2)b(y)$, say. This includes models for data such as angles and proportions, that are not well accommodated by the class of exponential dispersion models. Examples of such models are the von Mises distribution (see e.g. Mardia, 1972; Fisher, 1993) and the simplex distribution (Barndorff-Nielsen and Jørgensen, 1991). The von Mises distribution is a proper dispersion model with density

$$\frac{e^{\sigma^{-2}}}{2\pi I_0(\sigma^{-2})} \exp\left[-\frac{1}{2\sigma^2} 2\left\{1 - \cos(y-\mu)\right\}\right],\ y \in (-\pi, \pi),$$

where I_0 is the modified Bessel function of first type and order zero (Abramowitz and Stegun, 1972), $\mu \in (-\pi, \pi)$ is the circular mean and σ^2 is the dispersion parameter. The (standard) simplex distribution on the unit interval $(0, 1)$ with parameters $\mu \in (0, 1)$ and $\sigma^2 > 0$ is a proper dispersion model with probability density function

$$\left[2\pi\sigma^2\left\{y(1-y)\right\}^3\right]^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right\}.$$

This distribution is a useful alternative to the beta distribution for analysis of continuous proportions. Jørgensen (1997a,b) gives further details of dispersion models, and consider many useful examples of proper dispersion models.

We now define some important notation. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^\top$, $\mathbf{H}_i = \operatorname{diag}\left\{\dot{h}(\mathbf{x}_{i1}^\top\boldsymbol{\beta}), \ldots, \dot{h}(\mathbf{x}_{in_i}^\top\boldsymbol{\beta})\right\}$, where $\dot{h}$ indicates the derivative of h, $\mathbf{W}_i = \operatorname{diag}\{w_{i1}, \ldots, w_{in_i}\}$, $\mathbf{w}_i = (w_{i1}, \ldots, w_{in_i})^\top$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^\top$. Also, let $\mathbf{u}_i$ and $\dot{\mathbf{u}}_i$ be n_i-dimensional vectors with components

$$u_{it} = -\frac{\partial d}{2\partial\mu_{it}}(Y_{it};\mu_{it}) \text{ and } \dot{u}_{it} = -\frac{\partial^2 d}{2\partial\mu_{it}}(Y_{it};\mu_{it}),$$

respectively.

3 Estimating functions

Let us consider estimation of $\boldsymbol{\beta}$ based on estimating functions of the form $\boldsymbol{\Psi}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{\psi}_i(\mathbf{Y}_i, \boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\beta})$ where $\boldsymbol{\psi}_i$, $i = 1, \ldots, n$ are unbiased estimating functions. The Godambe information matrix for $\boldsymbol{\psi}_i$ is defined by $\mathbf{J}_i(\boldsymbol{\beta}) = \mathbf{S}_i(\boldsymbol{\beta})\mathbf{A}_i^{-1}(\boldsymbol{\beta})\mathbf{S}_i^\top(\boldsymbol{\beta})$, where

$$\mathbf{A}_i(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}}\left\{\boldsymbol{\psi}_i(\boldsymbol{\beta})\boldsymbol{\psi}_i^\top(\boldsymbol{\beta})\right\},\ \mathbf{S}_i(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}}\left\{\nabla_{\boldsymbol{\beta}}\boldsymbol{\psi}_i(\boldsymbol{\beta})\right\},$$

and $\nabla_{\boldsymbol{\theta}}$ is the gradient operator. Here we assume that ψ_i is regular, such that, in particular, $\mathbf{A}_i(\boldsymbol{\beta})$ and $\mathbf{S}_i(\boldsymbol{\beta})$ are finite and nonsingular and the operations of integration and differentiation with respect to $\boldsymbol{\beta}$ can be interchanged. For details see Jørgensen and Labouriau (1994) and McLeish and Small (1988).

By standard asymptotic theory for estimating functions, one may show that under certain regularity conditions, the sequence of roots, $\left\{\widehat{\boldsymbol{\beta}}_n\right\}_{n=1}^{\infty}$, associated with the estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\beta})$ is consistent for $\boldsymbol{\beta}$ and asymptotically normal. Specifically,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N\left\{0, \overline{\mathbf{J}}^{-1}(\boldsymbol{\beta})\right\}, \qquad \text{as } n \to \infty, \tag{2}$$

where

$$\overline{\mathbf{J}}(\boldsymbol{\beta}) = \lim_{n\to\infty} \left\{n^{-1}\sum_{i=1}^{n} \mathbf{S}_i(\boldsymbol{\beta})\right\} \left\{n^{-1}\sum_{i=1}^{n} \mathbf{A}_i(\boldsymbol{\beta})\right\}^{-1} \left\{n^{-1}\sum_{i=1}^{n} \mathbf{S}_i(\boldsymbol{\beta})\right\}^{\top}, \tag{3}$$

provided in particular that each of the averages involved converge.

In order to obtain the root of $\boldsymbol{\Psi}_n(\boldsymbol{\beta})$, one may use the following Newton scoring algorithm

$$\boldsymbol{\beta}^* = \boldsymbol{\beta} - \left\{\sum_{i=1}^{n} \mathbf{S}_i(\boldsymbol{\beta})\right\}^{-1} \boldsymbol{\Psi}(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ denotes the updated value of $\boldsymbol{\beta}$ (Jørgensen et al., 1996).

We now consider the class of *linear estimating functions* defined to be of the form

$$\sum_{i=1}^{n} \mathbf{Q}_i(\boldsymbol{\beta})\mathbf{b}_i(\boldsymbol{\beta}),$$

where $\mathbf{Q}_i(\boldsymbol{\beta})$ is a $p \times m$ matrix of constant weights and the $\mathbf{b}_i(\boldsymbol{\beta}) = \mathbf{b}_i(\mathbf{Y}_i; \boldsymbol{\beta})$ are independent m-dimensional zero mean vectors. Crowder (1987) proved that, in the class of all linear estimating functions with given $\mathbf{b}_i(\boldsymbol{\beta})$s and under regularity conditions, the optimum estimating function is obtained for the following choice of weight matrix,

$$\mathbf{Q}_i(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}}\left\{\nabla_{\boldsymbol{\beta}}\mathbf{b}_i(\boldsymbol{\beta})\right\} \mathrm{Cov}_{\boldsymbol{\beta}}^{-1}\left\{\mathbf{b}_i(\boldsymbol{\beta})\right\}.$$

This result will be used repeatedly in the following.

4 Position modelling

4.1 The GEE method

Consider first Liang and Zeger's GEE method in the case where the Y_{it} follow an exponential dispersion model. Here $\boldsymbol{\beta}$ is estimated by solving

$$\sum_{i=1}^{n} \mathbf{D}_i^{\top}\mathbf{C}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{4}$$

where $\mathbf{D}_i^\top = \mathbf{X}_i^\top \mathbf{H}_i$ and $\mathbf{C}_i$ is the working covariance matrix

$$\mathbf{C}_i = \sigma^2 \mathbf{W}_i^{1/2} \mathbf{V}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{V}_i^{1/2} \mathbf{W}_i^{1/2}, \tag{5}$$

where $\mathbf{V}_i = \operatorname{diag}\{V(\mu_{i1}), \ldots, V(\mu_{in_i})\}$, V is the variance function and $\mathbf{R}(\boldsymbol{\alpha})$ is the *working correlation matrix*, which depends on the vector parameter $\boldsymbol{\alpha}$.

Choosing $\mathbf{C}_i = \operatorname{Cov}(\mathbf{Y}_i)$, the function (4) is the optimum linear estimating equation in the exponential dispersion model case (Crowder, 1992).

4.2 Notation for dispersion models

The following notation facilitates the extension of Liang and Zeger's results. We define the *pseudo response vector* by

$$\widetilde{\mathbf{Y}}_i = \boldsymbol{\mu}_i + \mathbf{V}_{i\,\lambda w_i} \mathbf{u}_i,$$

where $\mathbf{V}_{i\,\lambda w_i}$ is a diagonal matrix with elements

$$V_{i\,\lambda w_{it}}(\mu_{it}) = \frac{\sigma^2}{w_{it} \operatorname{Var}(u_{it})}.$$

Note that the first Bartlett identity for the marginal distribution of Y_{it} gives

$$E(\dot{u}_{it}) = -\frac{1}{V_{i\,\lambda w_{it}}(\mu_{it})}. \tag{6}$$

For exponential dispersion models the function $V_{\lambda w_{it}}$ does not depend on λw_{it} and is equal to the usual variance function V which in turn makes $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i$. For general dispersion models, $\widetilde{\mathbf{Y}}_i$ depends on both $\boldsymbol{\mu}_i$ and σ^2, but the component $\widetilde{Y}_{it}$ has mean μ_{it} and variance

$$\operatorname{Var}\left(\widetilde{Y}_{it}\right) = \frac{\sigma^2}{w_{it}} V_{i\,\lambda w_{it}}(\mu_{it}). \tag{7}$$

4.3 Known covariance structure

We now extend Liang and Zeger's (1986) results by making second-moment assumptions about $\widetilde{\mathbf{Y}}$. Let us first consider the simple, but unrealistic case where the covariance matrix of $\widetilde{\mathbf{Y}}_i$ is known. Let $\mathbf{C}_i$, $i = 1, \ldots, n$, be given quadratic full-rank weight matrices and assume that σ^2 is known. Define an estimating function for $\boldsymbol{\beta}$ by $\Psi_n^C(\boldsymbol{\beta}) = \sum_{i=1}^n \psi_i^C(\boldsymbol{\beta})$, where

$$\psi_i^C(\boldsymbol{\beta}) = \mathbf{D}_i^\top \mathbf{C}_i^{-1} \left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right). \tag{8}$$

The components of the associated Godambe information matrix for $\boldsymbol{\beta}$ are given by

$$\mathbf{A}_i(\boldsymbol{\beta}) = \mathbf{D}_i^\top \mathbf{C}_i^{-1} \operatorname{Cov}(\widetilde{\mathbf{Y}}_i) \mathbf{C}_i^{-1} \mathbf{D}_i, \quad \mathbf{S}_i(\boldsymbol{\beta}) = -\mathbf{D}_i^\top \mathbf{C}_i^{-1} \mathbf{D}_i. \tag{9}$$

The asymptotic distribution of $\widehat{\boldsymbol{\beta}}_n$ is given by (2).

The optimum linear estimating equation is obtained by the choice $\mathbf{C}_i = \operatorname{Cov}(\widetilde{\mathbf{Y}}_i)$. This follows by using (6) and (7).

4.4 Unknown covariance structure

We now consider the more realistic case where the covariance structure of the data is not known. We assume a parametric structure $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\alpha})$ for the weight matrices, and consider two methods for estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Normally we assume a structure similar to (5), namely

$$\mathbf{C}_i = \sigma^2 \mathbf{W}_i^{1/2} \mathbf{V}_{i\lambda w_i}^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{V}_{i\lambda w_i}^{1/2} \mathbf{W}_i^{1/2},$$

which is consistent with the marginal variance structure of $\widetilde{\mathbf{Y}}_i$. However such an assumption is not necessary for the following results.

4.4.1 Substitution method

Consider an estimating function (not necessarily unbiased) given by

$$\Psi_n(\boldsymbol{\beta}) = \Psi_n(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}_n^*) = \sum_{i=1}^{n} \mathbf{D}_i^\top \mathbf{C}_i^{-1}(\widehat{\boldsymbol{\alpha}}_n^*) \left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \right), \tag{10}$$

where $\widehat{\boldsymbol{\alpha}}_n^* = \widehat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}) = \widehat{\boldsymbol{\alpha}}_n\{\boldsymbol{\beta}, \widehat{\sigma}_n^2(\boldsymbol{\beta})\}$. Here we assume the following conditions.

(a) Given $\boldsymbol{\beta}$ and σ^2, $\widehat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}, \sigma^2)$ is a $\sqrt{n}$- consistent estimator of $\boldsymbol{\alpha}$.

(b) Given $\boldsymbol{\beta}$, $\widehat{\sigma}_n^2(\boldsymbol{\beta})$ is a $\sqrt{n}$-consistent estimator of σ^2.

(c) The estimator $\widehat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}, \sigma^2)$ satisfies

$$\left| \frac{\partial \widehat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \right| \leq \mathbf{H}(\mathbf{Y}; \boldsymbol{\beta}) = O_p(1),$$

where $\mathbf{Y}$ is the data vector.

Under these and certain further regularity conditions, it can be proved that $\widehat{\boldsymbol{\beta}}_n$, the root of the function (10), is consistent and satisfies (2), where $\mathbf{S}_i$ and $\mathbf{A}_i$ are defined by (9) with $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\alpha})$. The proof of this result, presented in detail in Appendix A, is analogous to a result of Liang and Zeger (1986).

The asymptotic covariance matrix for $\widehat{\boldsymbol{\beta}}_n$ may be consistently estimated by the empirical sandwich estimator given by

$$\widehat{\overline{\mathbf{J}}}^{-1} = n \left\{ \sum_{i=1}^{n} \mathbf{S}_i \right\}^{-1} \sum_{i=1}^{n} \mathbf{D}_i^\top \mathbf{C}_i^{-1}(\widehat{\boldsymbol{\alpha}}_n^*) \left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \right) \left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \right)^\top \mathbf{C}_i^{-1}(\widehat{\boldsymbol{\alpha}}_n^*) \mathbf{D}_i \left\{ \sum_{i=1}^{k} \mathbf{S}_i \right\}^{-\top},$$

where $-\top$ denotes the inverse transpose matrix and all matrices are evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_n$.

The substitution method requires a consistent estimator for $\boldsymbol{\alpha}$ although $\widehat{\boldsymbol{\beta}}_n$ is consistent even if $\mathbf{C}_i(\boldsymbol{\alpha})$ does not correspond to the true covariance matrix of $\widetilde{\mathbf{Y}}_i$. However, the latter choice for $\mathbf{C}_i(\boldsymbol{\alpha})$ may lead to a more efficient estimator, as discussed in the case of exponential dispersion models by Liang et al. (1992), Fitzmaurice et al. (1993) and Albert and McShane (1995).

The dispersion parameter σ^2 may be estimated by the solution to

$$\sigma^2 = \frac{1}{\sum_{i=1}^{n} n_i - p} \sum_{i=1}^{n} \sum_{1}^{n_i} w_{it} \frac{\left(\widetilde{Y}_{it} - \mu_{it}\right)^2}{V_{i\,\lambda w_{it}}\left(\mu_{it}\right)}.$$

Note that σ^2 enters on the right-hand side of this equation via $\widetilde{Y}_{it}$ and $V_{i\,\lambda w_{it}}\left(\mu_{it}\right)$.

4.4.2 Estimating functions for covariance structure

We now consider an approach where the $\boldsymbol{\beta}$ estimating function is coupled with an estimation function for $\boldsymbol{\alpha}$.

In this section we let $\mathbf{C}_i = \mathrm{Cov}(\widetilde{\mathbf{Y}}_i)$ and $\boldsymbol{\sigma}_i$ is an $n_i\left(n_i+1\right)/2 \times 1$ vector with components $\sigma_{itk}(\boldsymbol{\alpha}) = \mathrm{Cov}(\widetilde{Y}_{it}, \widetilde{Y}_{ik})$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$. Define a vector $\mathbf{v}_i$ with components

$$\left(\widetilde{Y}_{it} - \mu_{it}\right)\left(\widetilde{Y}_{ik} - \mu_{ik}\right),$$

for $i = 1, \ldots, n$, $t = 1, \ldots, n_i$ and $k = 1, \ldots, t$, ordered in the same way as for $\boldsymbol{\sigma}_i$. The optimum linear estimating function for $\boldsymbol{\theta}$ is given by

$$\Psi_n^{\mathrm{cov}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{D}_i^\top & 0 \\ \frac{\partial \mathbf{v}_i}{\partial \boldsymbol{\beta}^\top} & \frac{\partial \mathbf{v}_i}{\partial \boldsymbol{\alpha}^\top} \end{bmatrix} \begin{bmatrix} \mathrm{Cov}(\widetilde{\mathbf{Y}}_i) & \mathrm{Cov}(\widetilde{\mathbf{Y}}_i, \mathbf{v}_i) \\ \mathrm{Cov}(\mathbf{v}_i, \widetilde{\mathbf{Y}}_i) & \mathrm{Cov}(\mathbf{v}_i) \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \\ \mathbf{v}_i - \boldsymbol{\sigma}_i \end{bmatrix}. \qquad (11)$$

These functions require knowledge of $\mathrm{Cov}(\mathbf{v}_i)$ and $\mathrm{Cov}(\widetilde{\mathbf{Y}}_i, \mathbf{v}_i)$, which in the exponential dispersion model case amounts to knowledge of the third and forth moments of Y_{it}.

In practice, $\mathrm{Cov}(\mathbf{v}_i)$ and $\mathrm{Cov}(\widetilde{\mathbf{Y}}_i, \mathbf{v}_i)$ must, in turn, be modelled as a function of further parameters, similar to what was done with $\mathbf{C}_i(\boldsymbol{\alpha})$ in Section 4.4.1. One may prove a result analogous to the result of that section, as in Prentice (1988) and Prentice and Zhao (1991).

To avoid making assumptions about $\mathrm{Cov}(\mathbf{v}_i)$ and $\mathrm{Cov}(\widetilde{\mathbf{Y}}_i, \mathbf{v}_i)$, one may use one of the following simplified versions of (11),

$$\Psi_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{D}_i^\top & 0 \\ 0 & \frac{\partial \mathbf{v}_i}{\partial \boldsymbol{\alpha}} \end{bmatrix} \begin{bmatrix} \mathrm{Cov}^{-1}(\widetilde{\mathbf{Y}}_i) & 0 \\ 0 & \mathrm{Cov}^{-1}(\mathbf{v}_i) \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \\ \mathbf{v}_i - \boldsymbol{\sigma}_i \end{bmatrix}, \qquad (12)$$

$$\Psi_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{D}_i^\top & 0 \\ 0 & \frac{\partial \mathbf{v}_i}{\partial \boldsymbol{\alpha}} \end{bmatrix} \begin{bmatrix} \mathrm{Cov}^{-1}(\widetilde{\mathbf{Y}}_i) & 0 \\ 0 & \mathbf{I}_{n_i} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i \\ \mathbf{v}_i - \boldsymbol{\sigma}_i \end{bmatrix}. \qquad (13)$$

Note that in both the above cases, the original estimating function (8) appears as the first component of each equation, which is not the case for (11).

The consistency of the above three $\boldsymbol{\theta}$-estimators requires $E(\mathbf{v}_i) = 0$, $i = 1, \ldots, n$, which in turn depends on the correct modelling of the covariance structure $\sigma_{ijk}(\boldsymbol{\alpha})$.

5 Dispersion and position modelling

We now extend the results of the previous section to the case of joint modelling of position and dispersion. The modelling of dispersion may be of interest in several practical situations, for instance, in the study of the direction of bird orientation, where the variability of the angle tends to decrease with time.

We consider a regression model for w_{it} given by

$$w_{it} = f\left(\mathbf{z}_{it}^{\top}\gamma\right),$$

where f is a twice differentiable, one-to-one inverse link function, γ is a $q \times 1$ parameter vector, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})^{\top}$, $i = 1, \ldots, n$ are matrices of covariates.

5.1 Naive method

Consider $\mathbf{C}_i$ as in Section 4 and define a set of estimating equation for $\boldsymbol{\theta} = (\beta^{\top}, \gamma^{\top}, \lambda)^{\top}$ by

$$\Phi_n^C(\boldsymbol{\theta}) = \sum_{i=1}^{n} \phi_i^C(\boldsymbol{\theta}) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{X}_i^{\top}\mathbf{H}_i\mathbf{C}_i^{-1}\left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right) \\ \mathbf{Z}_i^{\top}\mathbf{F}_i\left(\lambda\dot{\mathbf{c}}_i - \frac{\lambda}{2}\mathbf{d}_i\right) \\ \mathbf{w}_i^{\top}\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{w}_i^{\top}\mathbf{d}_i \end{bmatrix}, \tag{14}$$

where $\dot{\mathbf{c}}_i$ is an $n_i \times 1$ vector with components $\frac{\partial}{\partial x}c\left(y_{it}, x\right)|_{x=\lambda w_{it}}$, $\mathbf{d}_i$ is a $n_i \times 1$ vector with components $d\left(y_{it};\mu_{it}\right)$ and $\mathbf{F}_i = \operatorname{diag}\left\{\dot{f}\left(\mathbf{z}_{i1}^{\top}\gamma\right), \ldots, \dot{f}\left(\mathbf{z}_{it}^{\top}\gamma\right)\right\}^{\top}$.

If the $\mathbf{C}_i$s are known, then under general regularity conditions, the sequence of roots of (14) is consistent and asymptotically normal as in (2) and (3), with

$$\mathbf{S}_i(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{S}_{11i} & 0 & 0 \\ 0 & \mathbf{S}_{22i} & \mathbf{S}_{23i} \\ 0 & \mathbf{S}_{32i} & \mathbf{S}_{33i} \end{pmatrix} \text{ and } \mathbf{A}_i(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A}_{11i} & \mathbf{A}_{12i} & \mathbf{A}_{13i} \\ \mathbf{A}_{21i} & \mathbf{A}_{22i} & \mathbf{A}_{23i} \\ \mathbf{A}_{31i} & \mathbf{A}_{32i} & \mathbf{A}_{33i} \end{pmatrix}, \tag{15}$$

whose components may be found in Appendix B.

These expressions become considerably easier when the model $\mathrm{DM}(\mu;\sigma^2)$ is a proper dispersion model, where $\mathbf{c}_i$ is constant.

One may model $\mathbf{C}_i$ as a function of an unknown parameter vector $\boldsymbol{\alpha}$, proceeding analogously to Section 4.4.

Let $\widehat{\boldsymbol{\theta}}_n$ be the root of (14). Under regularity conditions and assuming that $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ is a $n^{1/2}$-consistent estimator of $\boldsymbol{\alpha}$ for given $\boldsymbol{\theta}$, we obtain

$$n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}\left[0, \lim_{n\to\infty} n\left\{\sum_{i=1}^{n}\mathbf{S}_i(\boldsymbol{\theta})\right\}^{-1}\left\{\sum_{i=1}^{n}\mathbf{A}_i(\boldsymbol{\theta})\right\}\left\{\sum_{i=1}^{k}\mathbf{S}_i(\boldsymbol{\theta})\right\}^{-\top}\right],$$

when $n \to \infty$, $\mathbf{S}_i$ and $\mathbf{A}_i$ as in (15). The proof of this theorem is similar to the analogous result of Section 4.4 and of Prentice and Zhao (1991).

5.2 Optimum linear estimating function

The estimating function Φ_n^C is not optimum in the class of linear estimating functions. Let us now consider the optimum linear estimating function followed by an intermediate case.

The optimum linear estimating function is given by

$$\Phi_n^*(\theta) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{X}_i^\top \mathbf{H}_i & 0 & 0 \\ 0 & \mathbf{Z}_i^\top \mathbf{F}_i \mathbf{K}_i & \lambda \mathbf{W}_i E(\ddot{\mathbf{c}}_i) \\ 0 & \lambda E\left(\ddot{\mathbf{c}}_i^\top\right) \mathbf{W}_i & \mathbf{1}\left(\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{d}_i\right) \end{bmatrix} \mathrm{Cov}^{-1}(\mathbf{s}_i)\,\mathbf{s}_i,$$

where $\mathbf{s}_i^\top = \left\{\left(\widetilde{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top, \left(\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{d}_i\right)^\top, \left(\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{d}_i\right)^\top \mathbf{1}\right\}$ and $\ddot{\mathbf{c}}_i$ is a $n_i \times 1$ vector with components $\frac{\partial^2 c}{\partial x^2}(y_{it}; x)\,|_{x=\lambda w_{it}}$. If $\mathrm{Cov}(\mathbf{s}_i)$ is unknown, this function requires the estimation of $(2n_i + 1)(2n_i + 3)/2$ nuisance parameters corresponding the components of this matrix.

A simpler set of equations is obtained as follows

$$\Phi_n^o(\theta) = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{X}_i^\top \mathbf{H}_i & 0 & 0 \\ 0 & \mathbf{Z}_i^\top \mathbf{F}_i \mathbf{K}_i & \lambda \mathbf{W}_i E(\ddot{\mathbf{c}}_i) \\ 0 & \lambda E\left(\ddot{\mathbf{c}}_i^\top\right) \mathbf{W}_i & \left(\mathbf{c}_i - \frac{1}{2}\mathbf{d}_i\right) \end{bmatrix} \mathbf{G}_i \mathbf{s}_i. \tag{16}$$

where $\mathrm{Cov}^{-1}(\mathbf{s}_i)$ is replaced by the block diagonal matrix

$$\mathbf{G}_i = \begin{bmatrix} \mathrm{Cov}^{-1}\left(\widetilde{\mathbf{Y}}_i\right) & 0 & 0 \\ 0 & \mathrm{Cov}^{-1}\left(\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{d}_i\right) & 0 \\ 0 & 0 & \mathrm{Var}^{-1}\left(\dot{\mathbf{c}}_i - \frac{1}{2}\mathbf{d}_i\right)^\top \mathbf{1} \end{bmatrix}.$$

6 Simulations

The following simulation study concerns the behaviour of position modelling for directional data. We consider a circular random variable Y_{it} with circular mean

$$\mu_{it} = \beta_0 + 2\arctan(\beta x_{it}), \; i = 1, \dots, n, \; j = 1, \dots, n_i,$$

where $\beta_0 = 1$, $\beta = 0.2$ and the x_{it} were generated from a standard normal distribution.

Let $\mathbf{Z}_i$ be a $n_i \times 1$ normal random vector with components Z_{it}, mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\omega^2\mathbf{R}$, where $\mathbf{R}$ is a correlation matrix. Let $\mathbf{Y}_i$ be a circular random vector, with components $Y_{it} = Z_{it} \bmod 2\pi$, so that the Y_{it} follow correlated wrapped normal distributions with circular means μ_{it} and dispersion parameters ω^2 (cf. Fisher, 1993, p. 55). Many authors, including Mardia (1972, pp. 66–67), Fisher (1993, p. 55), Breckling (1989, pp. 138–140), Collet and Lewis (1981), Watson (1982) and Stephen (1963), have found that the wrapped normal distribution with dispersion parameter ω^2 may be closely approximated by a von Mises distribution with concentration parameter λ given by the relation

$$A_1(\lambda) = \frac{I_1(\lambda)}{I_0(\lambda)} = \exp\left(-\frac{1}{2}\omega^2\right), \tag{17}$$

Table 1: Mean square error for independence function relative to substitution method.

		n_i					
		5			12		
		ρ			ρ		
n	λ	0	0.5	0.9	0	0.5	0.9
20	7	0.71	1.12	5.45	0.31	0.52	2.62
	2	0.72	1.05	4.65	0.31	0.50	2.31
	1	0.70	0.89	0.97	0.33	0.41	1.38
50	7	0.89	1.50	7.25	0.71	1.22	6.38
	2	0.91	1.38	5.99	0.72	1.12	5.60
	1	0.90	1.13	4.07	0.73	0.96	0.96
100	7	0.94	1.48	7.69	0.88	1.43	8.00
	2	0.96	1.39	6.75	0.87	1.32	6.15
	1	0.96	1.18	4.45	0.88	1.10	4.07

where $I_r(\cdot)$ is the modified Bessel function of order r (see Abramowitz and Stegun, 1972). Note that the pseudo-response variables are given by

$$\mu_{it} + A_1(\lambda)\sin(Y_{it} - \mu_{it})\,.$$

It is worth noting that the marginal distributions do not have to be von Mises for the asymptotic results developed above to be valid for directional data. It is enough that $E\{\sin(Y_{it} - \mu_{it})\} = 0$, which is satisfied by any circular distribution.

In the first set of simulations we compare the substitution method with estimation based on the von Mises score function, assuming the independence correlation structure, the latter called the independence function from now on. We simulated a thousand samples for each combination of $n = 20, 50, 100$, $n_i = 5, 12$, $\lambda = 7, 2, 1$ and $\rho = 0, 0.5, 0.9$, estimating the parameters by the Newton scoring algorithm. No special structure was considered for the correlation matrix $\mathbf{R}$, and its elements were estimated by the Pearson correlation.

Table 1 shows the ratio of the mean square error for estimation with the independence function relative to that for the substitution method. We note that the performance of the substitution method increases as a function of ρ and n, and decreases as a function of n_i. The behaviour of the substitution method is particularly bad when $n = 20$ and $n_i = 12$, where the number of nuisance parameters is large relative to n. It seems reasonable to recommend the substitution method when n is large relative to n_i and ρ is large.

The second set of simulations compares estimation based on the independence function with maximum likelihood estimation. Due to computational limitations, the covariance matrix of $\mathbf{Z}_i$ was considered known and only 200 simulations were conducted for each combination of $n = 20, 50$, $n_i = 5, 12$, $\lambda = 7, 2$ and $\rho = 0, 0.5, 0.9$.

Table 2 shows the observed mean absolute estimation error for the maximum likelihood method (MLE), the independence function, and the substitution method.

Table 2: Mean absolute estimation error times 100 for MLE, independence function and substitution method.

			n_i					
			5			12		
			ρ			ρ		
n	λ	Method	0	0.5	0.9	0	0.5	0.9
20	7	MLE	1.78	1.71	0.65	1.28	0.89	0.38
		Independence	1.80	1.77	1.80	1.16	1.16	1.16
		Substitution	2.13	1.71	0.76	2.02	1.60	0.70
	2	MLE	3.80	3.11	1.39	2.48	1.93	0.91
		Independence	4.03	3.97	4.11	2.62	2.65	2.63
		Substitution	4.72	3.91	1.85	4.51	3.68	1.70
50	7	MLE	1.18	0.99	0.35	0.86	0.58	0.25
		Independence	1.09	1.09	1.09	0.72	0.73	0.72
		Substitution	1.16	0.91	0.40	0.86	0.66	0.28
	2	MLE	2.28	1.81	0.80	1.60	1.27	0.53
		Independence	2.41	2.43	2.46	1.59	1.64	1.63
		Substitution	2.53	2.08	0.96	1.87	1.55	0.70

The mean absolute errors are almost constant as a function of ρ for the independence function, but tend to decreases with ρ for the other methods. For small to moderate ρ the performance of the independence function is similar to maximum likelihood, whereas for $\rho = 0.9$ the substitution method performs somewhat similar to maximum likelihood.

We conclude that good performance of the substitution method for the circular regression model requires either a high correlation or a large sample size.

Regarding the convergence of the Newton scoring algorithm for the substitution method, there were no problems for large n and λ, but for $\lambda = 1$, $n = 20$ or 50, $\rho = 0.5$ or 0.9 and $n_i = 5$, up to 5% of the cases failed to converge. For $n_i = 12$ $n = 20$, $\lambda = 1$ and $\rho = 0.9$ the percentage of failed cases increased to about 20%. These problems may be due to the high number of nuisance parameters compared with n_i, similar to the findings of Lipsitz et al. (1994) for the original GEE method.

7 Bird orientation data

We consider data from a study of the orientation mechanism of pigeons conducted in 1982 in the state of Ceará, Brazil (Ranvaud et al. 1983). The pigeons were moved in a lightproof cage from Fortaleza to the release point 300km away near the city of Camocim. The responses were the angular differences between the pigeons' vanishing position and their positions at

Table 3: Circular means and concentration parameters for the response vectors.

Release point	Time of day	μ_{it} 30s	μ_{it} 60s	μ_{it} 90s	λ 30s	λ 60s	λ 90s	n
I	morning	10.5	12.5	10.1	2.06	2.27	3.89	37
	noon	19.9	21.2	12.7	1.23	1.45	1.84	30
	afternoon	16.9	13.6	3.4	1.75	2.68	4.29	28
II	morning	−3.5	−7.2	−5.0	1.36	2.09	3.38	28
	noon	−24.1	−9.5	−2.3	1.15	1.67	2.83	28
	afternoon	11.1	9.1	7.3	2.21	3.37	6.81	25

($t \times 30$)

30, 60 and 90 seconds after release. We investigate the effects of release point (I and II) and time of day of release (morning, noon, afternoon).

Let Y_{it} be the angle of pigeon i at $t \times 30$ seconds after release ($t = 1, 2, 3$), with circular mean μ_{it} and concentration parameter λ_{it}. Suppose that the pigeons are independent, that the dependence structure is the same for each combination of release point and time of day and that μ_{it} and λ_{it} depend only on release point, time of day and t. Table 3 shows the corresponding observed circular means and concentration parameters.

The full model used for these data was

$$\mu_{it} = 2 \arctan \{\eta_{0it} + \eta_{1it}(t-2)\}$$
$$\lambda_{it} = \exp \{\eta_{2it} + \eta_{3it}(t-2)\},$$

$$\eta_{mit} = \beta_m + \beta_{mo} x_{1it} + \beta_{ma} x_{2it} + \beta_{ml} x_{3it} + \beta_{mol} x_{1it} x_{3it} + \beta_{mal} x_{2it} x_{3it},$$

for $m = 0, 1, 2, 3$, where x_{1it}, x_{2it} and x_{3it} are indicator functions with value one when pigeon i is released at noon, afternoon and release point two, respectively.

We assume an unstructured correlation matrix for the s_i. The first three components of s_i are given by $\mathcal{A}_1(\lambda_{it}) \sin(Y_{it} - \mu_{it})$ and the last three by $\mathcal{A}_1(\lambda_{it}) \cos(Y_{it} - \mu_{it})$. Due to the relatively small sample sizes compared with the number of nuisance parameters we chose the estimating function (16).

After fitting the full model, we tested if all parameters related to μ_{it}, except β_0 (constant), are zero. We obtained a Wald test of 15.75 on 11 degrees of freedom, which gave p-value of 0.151 based on the asymptotic a chi-squares distribution, so these parameters were hence removed from the model.

Turning now to the concentration parameters, we tested if all coefficients related to λ_{it}, except for β_2 (constant), β_{2o} (noon indicator) and β_3 (trend) were different from zero. The Wald test was 10.24 on 9 degrees of freedom, giving a p-value of 0.331. Removing these parameters gave the final model, whose estimates are shown in Tables 4 and 5.

Table 4: Parameter estimates for final model.

	Estimate	s.e.	t	p-value
β_0	0.03	0.02	2.06	0.002
β_2	1.28	0.19	6.88	< 0.001
β_{2o}	-0.28	0.27	-1.06	0.013
β_3	0.52	0.07	7.71	< 0.001

Table 5: Estimated concentration parameters for final model.

	$t \times 30$		
Time of day	30s	60s	90s
morning/afternoon	2.02	2.87	4.09
noon	1.23	1.75	2.49

According to the final model the circular mean is constant, with estimated value 5.1°. The concentration parameter increases with t and is smaller when the pigeons are released at noon.

Acknowledgments

We are grateful to Professor Ronald Ranvaud of the Institute of Biomedical Science, University of São Paulo, for permission to use the data analysed in Section 7. The research was supported by the National Science and Engineering Council of Canada, CNPq – Brazil and FAPESP – Brazil.

A Proof of results for substitution method

To show the results of Section 4.4, we first prove some lemmas. Let $\mathcal{H}$ denote the conditions (a)–(c) of Section 4.4.

Lemma 1 *Under $\mathcal{H}$ and assuming that $\widehat{\alpha}_n^*$ is differentiable with respect to σ^2, we have*

$$\sqrt{n}(\widehat{\alpha}_n^* - \alpha) = O_p(1).$$

Proof: We have

$$\sqrt{n}(\widehat{\alpha}_n^* - \alpha) = \sqrt{n}\{\widehat{\alpha}_n(\beta, \widehat{\sigma}_n^2(\beta)) - \widehat{\alpha}_n(\beta, \sigma^2) + \widehat{\alpha}_n(\beta, \sigma^2) - \alpha\},$$

and by condition (a), $\sqrt{n}\,\{\widehat{\alpha}_n(\beta,\sigma^2)-\alpha)\} = O_p(n^{-1/2})$. On the other hand, by a Taylor-expansion of $\widehat{\alpha}_n(\beta,\sigma^2)$ around $\hat{\sigma}^2_n$, and using condition (b), it follows that

$$\widehat{\alpha}_n\left(\beta,\sigma^2\right) = \widehat{\alpha}_n\left(\beta,\hat{\sigma}^2_n\right) + \frac{\partial\widehat{\alpha}_n}{\partial\sigma^2}\left(\beta,\hat{\sigma}^2_n\right)\left(\sigma^2-\hat{\sigma}^2_n\right) + O_p\left(n^{-1/2}\right).$$

As n tending to infinity, applying conditions (b) and (c), it follows that

$$\widehat{\alpha}_n\left(\beta,\hat{\sigma}^2_n(\beta)\right) - \widehat{\alpha}_n\left(\beta,\sigma^2\right) = -\frac{\partial\widehat{\alpha}_n}{\partial\sigma^2}\left(\beta,\hat{\sigma}^2_n\right)\left(\sigma^2-\hat{\sigma}^2_n\right) + O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right).$$

Hence

$$\sqrt{n}\,(\alpha^*_n-\alpha) = O_p(1). \quad \square$$

Lemma 2 *Under $\mathcal{H}$, assuming that $\partial\Psi_n(\beta,\alpha)/\partial\alpha$ satisfies Markov's condition and that $\max_{1\leq i\leq n}\max_{1\leq t\leq n_i} E\mid\Psi_{it}\mid^{2+\delta}<\infty$ for some $0<\delta<1$, where Ψ_{it} is the t-th component of the vector Ψ_i, we have*

$$\frac{1}{\sqrt{n}}\Psi_n\left(\beta,\hat{\alpha}^*_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0,\mathrm{A}) \text{ as } n\to\infty,$$

where $\mathrm{A} = \frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^n \mathrm{A}_i(\beta)$, as in (9).

Proof: Considering $\frac{1}{\sqrt{n}}\Psi_n(\beta,\widehat{\alpha}^*_n)$ as a function of α^* and making a Taylor-expansion in the neighbourhood of α, it follows that

$$\frac{1}{\sqrt{n}}\Psi_n(\beta,\widehat{\alpha}^*_n) = \frac{1}{\sqrt{n}}\Psi_n(\beta,\alpha) + \frac{1}{n}\frac{\partial\Psi_n}{\partial\alpha}(\beta,\alpha)\sqrt{n}(\widehat{\alpha}^*_n-\alpha) + \frac{1}{\sqrt{n}}O_p\left(n^{-1/2}\right). \tag{18}$$

Note that $\frac{1}{n}\frac{\partial\Psi_n}{\partial\alpha}(\beta,\alpha) = \frac{1}{n}\sum_{i=1}^n\frac{\partial\psi_i}{\partial\alpha}(\beta,\alpha)$. Since $\frac{\partial\Psi_n}{\partial\alpha}(\beta,\alpha)$ depends on the response variable via a linear combination of independent zero mean vectors, Markov's weak law of large numbers assures that $\frac{1}{n}\frac{\partial\Psi_n}{\partial\alpha}(\beta,\alpha) = O_p(1)$. Applying this result and Lemma 1 to (18), it follows that $\frac{1}{\sqrt{n}}\Psi_n(\beta,\widehat{\alpha}^*_n)$ is asymptotically equivalent to $\frac{1}{\sqrt{n}}\Psi_n(\beta,\alpha)$ in distribution. Since $\max_{1\leq i\leq n}\max_{1\leq t\leq n_i} E\mid\Psi_{it}\mid^{2+\delta}<\infty$, Lyapunov's theorem assures that

$$\frac{1}{\sqrt{n}}\Psi_n(\beta,\alpha) \xrightarrow{\mathcal{D}} \mathcal{N}_p\left\{0;\mathrm{A}(\beta)\right\} \text{ as } n\to\infty. \tag{19}$$

The result now follows. $\square$

Lemma 3 *Let $\{\mathrm{T}_n\}_{n\geq 1}$ be a sequence of estimators such that $\mathrm{T}_n \xrightarrow{\mathcal{P}} \theta$ as $n\to\infty$, $\{g_n(\theta)\}_{n\geq 1}$ a sequence of measurable twice differentiable functions such that $g_n(\theta)\xrightarrow{\mathcal{P}} g(\theta)$ as $n\to\infty$, and $\dot{g}_n(\theta) = \partial g_n(\theta)/\partial\theta = O_p(1)$, then $g_n(\mathrm{T}_n)\xrightarrow{\mathcal{P}} g(\theta)$.*

Proof: A Taylor-expansion of $g_n(\mathrm{T}_n)$ near $\boldsymbol{\theta}$ gives

$$g_n(\mathrm{T}_n) = g_n(\boldsymbol{\theta}) + \dot{g}_n(\boldsymbol{\theta})(\mathrm{T}_n - \boldsymbol{\theta}) + O(\|\mathrm{T}_n - \boldsymbol{\theta}\|) = g_n(\boldsymbol{\theta}) + o_p(1).$$

Since $g_n(\boldsymbol{\theta}) \xrightarrow{\mathcal{P}} g(\boldsymbol{\theta})$, the result follows. □

Lemma 4 *Let $\dot{\Psi}_n(\beta, \hat{\alpha}_n^*) = \frac{\partial \Psi_n}{\partial \beta}(\beta, \hat{\alpha}_n^*)$, then under $\mathcal{H}$ and the assumptions of Lemma 3 applied to the components of $\dot{\Psi}_n(\beta, \hat{\alpha}_n^*) = g_n(\hat{\alpha}_n^*)$, it follows that*

$$\frac{1}{n}\dot{\Psi}_n(\beta, \hat{\alpha}_n^*) \xrightarrow{\mathcal{P}} \mathrm{S} \text{ as } n \to \infty,$$

where $\mathrm{S} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{S}_i(\beta)$ *is as in (9).*

Proof: Straightforward.

Proof of results for substitution method: Using the notation and definitions of Lemmas 1, 2 and 4, one can Taylor-expand $\Psi_n(\beta, \hat{\alpha}_n^*(\beta))$ as follows:

$$\Psi_n(\beta, \hat{\alpha}_n^*) = \Psi_n(\beta_0, \hat{\alpha}_n^*) + \dot{\Psi}_n(\beta_0, \hat{\alpha}_n^*)(\beta - \beta_0) + O(\| \beta - \beta_0 \|). \tag{20}$$

Evaluating the function at the point $\hat{\beta}_n$, and using the regularity conditions, it follows that

$$\Psi_n(\beta, \hat{\alpha}_n^*) + \dot{\Psi}_n(\beta, \hat{\alpha}_n^*)(\hat{\beta}_n - \beta) + O_p(1) = 0.$$

Hence, for n tending to infinity we have that

$$\sqrt{n}(\hat{\beta}_n - \beta) = \left\{\frac{1}{n}\dot{\Psi}_n(\beta_0)\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\Psi_n(\beta_0)\right\} + o_p(1). \tag{21}$$

The result now follows by applying Lemmas 2 and 4. □

B The components of (15)

$$\mathrm{S}_{11i} = \mathbf{D}_i^\top \mathbf{C}_i^{-1} E\{\mathrm{diag}(\dot{\mathrm{u}}_i)\}\mathbf{D}_i,$$

$$\mathrm{S}_{23i} = \mathrm{S}_{32i}^\top = \mathbf{Z}_i^\top \mathbf{F}_i \left[-\sigma^2 E(\mathbf{P}_{2i}) - \mathbf{W}_i^{-2} E\left\{\mathrm{diag}^{-1}(\mathbf{W}_i)\dot{\mathbf{W}}_i\right\} + \frac{1}{2\sigma^4}E(\mathrm{d}_i)\right],$$

$$\mathbf{P}_{2i} = \mathbf{W}_i^{-3}\delta\left[\left\{\mathrm{diag}(\ddot{\mathbf{W}}_i)\mathrm{diag}(\mathbf{W}_i) - \mathrm{diag}^2(\dot{\mathbf{W}}_i)\right\}\mathrm{diag}^{-2}(\mathbf{W}_i)\right],$$

$$\mathrm{S}_{22i} = \sigma^2 \mathbf{Z}_i^\top \mathbf{F}_i \mathbf{W}_i^{-1}\left\{\sigma^2 E(\mathbf{P}_{2i}) + 2E(\mathbf{P}_{1i})\right\}\mathbf{F}_i \mathbf{Z}_i,$$

$$\mathrm{S}_{33i} = \sum_{j=1}^{t}\left[\frac{1}{w_{it}^2}E\left(\frac{\ddot{a}_{it}a_{it} - \dot{a}_{it}^2}{a_{it}^2}\right) - \frac{w_{it}}{\sigma^6}E\left\{d(y_{ij}; \mu_{ij})\right\}\right],$$

$$A_{11i} = D_i^\top C_i^{-1} \mathrm{Cov}(u_i) C_i^{-1} D_i,$$
$$A_{12i} = A_{21i}^\top = -D_i^\top C_i^{-1} E\left\{u_i(\sigma^2 P_{i1} + \frac{1}{2\sigma^2} d_i)^\top\right\} F_i Z_i,$$
$$A_{13i} = A_{31i} = D_i^\top C_i^{-1} E\left(u_i s_i\right),$$
$$A_{22i} = Z_i^\top F_i E\left\{(\sigma^2 P_{1i} + \frac{1}{2\sigma^2} d_i)(\sigma^2 P_{1i} + \frac{1}{2\sigma^2} d_i)^\top\right\} F_i Z_i,$$
$$A_{23i} = A_{32i}^\top = -Z_i^\top F_i E\left\{(\sigma^2 P_i^{-2} + \frac{1}{2\sigma^2} d_i) s_i\right\}$$
$$A_{33i} = E(s_i^2).$$

References

[1] ABRAMOWITZ, M. AND STEGUN, I.A. (1972). *Handbook of Mathematical Functions.* New York: Dover.

[2] ALBERT, P.S. AND MCSHANE, L.M. (1995). A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimage data. *Biometrics* 50, 627–638.

[3] BARNDORFF-NIELSEN, O.E. AND JØRGENSEN, B. (1991). Some parametric models on the simplex. *J. Multivariate Anal.* 39, 106–116.

[4] BRECKLING, J. (1989). *Analysis of Directional Time Series: Application to Wind Speed and Direction.* Berlin: Springer-Verlag.

[5] COLLETT, D. AND LEWIS, T. (1981). Discriminating between the von Mises and Wrapped Normal distributions. *Austr. J. Statist.* 23, 73–79.

[6] COLOGNE, J.B., CARTES, R.L., FUJITA, S. AND BAN, S. (1993). Applications of generalized estimating equations to a study of *in vitro* radiation sensitivity. *Biometrics* 49, 927–934.

[7] CROWDER, M. (1987). On linear and quadratic estimating function. *Biometrika* 74, 591–597.

[8] CROWDER, M. (1992). Comment on the paper Liang, Zeger and Qaquish (1992). Multivariate regression analysis for categorical data. *J. Roy. Statist. Soc.* Ser. B 54, 26-27.

[9] DIGGLE, P.J., LIANG, K-Y AND ZEGER, S.L. (1994). *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

[10] FISHER, N.I. (1993). *Statistical Analysis of Circular Data.* Cambridge: Cambridge University Press.

[11] FITZMAURICE, G.M., LAIRD, N.M. AND ROTNITZKY, A.G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statist. Sci.* **8**, 284–309.

[12] JØRGENSEN, B. (1997a). Proper dispersion models. *Brazilian J. Probab. Statist.* (to appear).

[13] JØRGENSEN, B. (1997b). *The Theory of Dispersion Models.* London: Chapman & Hall.

[14] JØRGENSEN, B. AND LABOURIAU, R.S. (1994). *Exponential Families and Theoretical Inference.* Lecture Notes, Department of Statistics, University of British Columbia.

[15] JØRGENSEN, B., LUNDBYE-CHRISTENSEN, S., SONG, P.X.-K. AND SUN, L. (1996). State-space models for multivariate longitudinal data of mixed types. *Canad. J. Statist.* **24**, 385–402.

[16] KENWARD, M.G., LESAFFRE, E. AND MOLENBERGHS, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945–953.

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

[18] LIANG, K.-Y. AND HANFELT, J. (1994). On the use of quasi-likelihood method in teratological experiments. *Biometrics* **50**, 872-880.

[19] LIANG, K.-Y. AND ZEGER, S.L. (1986). Longitudinal analysis using generalized linear models. *Biometrika* **73**, 13–22.

[20] LIANG, K.-Y., ZEGER, S.L. AND QAQISH, B. (1992). Multivariate regression analysis for categorical data (with discussion). *J. Roy. Statist. Soc.* Ser. B **54**, 3–40.

[21] LIPSITZ, S.R., FITZMAURICE, G.M., ORAV, R.J., AND LAIRD, N.M. (1994). Performance of generalized estimating equations in practical situations. *Biometrics* **50**, 270–278.

[22] MARDIA, K.V. (1972). *Statistics of Directional Data.* London: Academic Press.

[23] MCLEISH, D.L. AND SMALL, C.G. (1988). *The Theory and Applications of Statistical Inference Functions. Lecture Notes in Statistics*, **44**. New York, Springer-Verlag.

[24] MILLER, M.E., DAVIS, C.S. AND LANDIS, J.R. (1993). The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics* **49**, 1033–1044.

[25] PRENTICE, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033–1048.

[26] PRENTICE, R.L. AND ZHAO, L.P. (1991). Estimating equation for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 47, 825–839.

[27] RANVAUD, R., SCHMIDT-KOENIG, K. KIEPENHEUER, J. AND GASPAROTTO, O.C. (1983) Initial orientation of homing pigeons at the magnetic equator with and without sun compass. *Behavioural Ecology and Sociobiology* 14, 77-79.

[28] STEPHENS, M.A. (1963). Random walk on a circle. *Biometrika* 50, 385–390.

[29] WATSON, G.S. (1982). Distributions on the circle and sphere. In Gani, J. and Hannan, E.J. (eds.) *Essays in Statistical Science.* Sheffield, Applied Probability Trust pp. 265–80. (Journal of Applied Probability, special volume 19A).

[30] WEDDERBURN, R.W.M. (1974). Quasi-likelihood function, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439–447.

[31] WHITTEMORE, A.S. AND GONG, G. (1994). Segregation analysis of case-control data using generalized estimating equations. *Biometrics* 50, 1073–1087.

[32] ZEGER, S.L. AND LIANG, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42, 121–130.

[33] ZEGER, S.L., LIANG, K.-Y. AND ALBERT, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060.

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

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9701 - **BOFARINE, H.; ARELLANO-VALE, R.B.** Weak nondifferential measurement error models. IME-USP, 1997. 12p. (RT-MAE-9701)

9702 - **FERRARI, S.L.P.; CORDEIRO, G. M.; CRIBARI-NETO, F.** Higher Order Asymptotic Refinements for Score Tests in Proper Dispersion Models. IME-USP, 1997. 14p. (RT-MAE-9702)

9703 - **DOREA, C.; GALVES, A.; KIRA, E.; ALENCAR, A. P.** Markovian modeling of the stress contours of Brazilian and European Portuguese. IME-USP, 1997. 11p. (RT-MAE-9703)

9704 - **FONTES, L.R.G.; ISOPI, M.; KOHAYAKAWA, Y.; PICCO, P.** The Spectral Gap of the REM under Metropolis Dynamics. IME-USP, 1997. 24p. (RT-MAE-9704)

9705 - **GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.A.** Estimation in weibull regression model with measurement error.IME-USP, 1997. 17p. (RT-MAE-9705)

9706 - **GALVES, A.; GUIOL, H.** Relaxation time of the one-dimensional symmetric zero range process with constant rate. IME-USP, 1997. 10p. (RT-MAE-9706)

9707 - **GUIOL, H.** A note about a burton keane's theorem. IME-USP, 1997. 7p. (RT-MAE-9707)

9708 - **ARELLANO-VALLE,R.B.; FERRARI,S.L.P.; CRIBARI-NETO,F.** Bartlett and barblett-type corrections for testing linear restrictions. IME-USP, 1997. 5p. (RT-MAE-9708)

9709 - **PAULA, G.A.** One-Sided Tests in Dose-Response Models. IME-USP, 1997. 29p. (RT-MAE-9709)

9710 - **BRESSAUD, X.** Subshift on an infinite alphabet. IME-USP, 1997. 34p. (RT-MAE-9710)

9711 - **YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H.** Stochastic Monotonicity properties of bayes estimation of the population size for capture-recapture data. IME-USP. 1997. 12p. (RT-MAE-9711)

9712 - **GUIMENEZ,P.;COLOSIMO,E.A.; BOLFARINE,H.** Asymptotic relative efficiency of wald tests in measurement error models. IME-USP. 1997. 14p. (RT-MAE-9712)

9713 - **SCHONMANN, R.H.; TANAKA, N.I.** Lack of monotonicity in ferromagnetic ising model phase diagrams. IME-UPS. 1997. 12p. (RT-MAE-9713)

9714 - **LEITE, J.G.; TORRES, V.H.S.; TIWARI, R.C.; ZALKIKAR, J.** Bayes estimation of dirichlet process parameter. IME-USP. 1997. 21p. (RT-MAE-9714)

9715 - **GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.** Hypotheses testing based on a corrected score function for errors-in-variables models. 1997. 16p. (RT-MAE-9715)

9716 - **PAULA, G.A.; BOLFARINE, H.** Some results on the slope of the linear regression model for the analysis of pretest-posttest data. 1997. 08p. (RT-MAE-9716)

9717 - **AUBIN, E.C.Q.; CORDEIRO, G.M.** Some adjusted likelihood ratio tests for heteroscedastic regression models. 1997. 19p. (RT-MAE-9717)

9718 - **ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a one-stage dose-response model using the minimum sum absolute errors estimators. 1997. 09p. (RT-MAE-9718)

9719 - **ANDRÉ, C.D.S.; NARULA, S.C.** Statistical inference for the parameters of a two-stage dose-response model using the minimum sum of absolute errors estimators. 1997. 15p. (RT-MAE-9719)

9720 - **COLLET, P.; GALVES, A.; SCHMITT, B.** Fluctuations of repetition time for gibbsian sources. 1997. 08p. (RT-MAE-9720)

9721 - **CORDEIRO, G.M.; FERRARI, S.L.P.; CYSNEIROS, A.H.M.A** A formula to improve score test statistics. 1997. 10p. (RT-MAE-9721)

9722 - **ZUAZOLA, P.I.; GASCO-CAMPOS, L.; BOLFARINE, H.** Pearson type II measurement error models. 1997. 24p. (RT-MAE-9722)

9723 - **BUENO, V.C.** A note on asymptotics reliability of a linearly connected system. 1997. 07p. (RT-MAE-9723)

9724 - **NARULA, S.C.; SALDIVA, P.H.N; ANDRÉ, C.D.S.; ELIAN, S.N.; FERREIRA, A.F.; CAPELOZZI, V.** The minimum sum of absolute errors regression: an overview. 1997. 22p. (RT-MAE-9724)

9725 - **BOLFARINE, H.; LIMA, C.R.O.P.; SANDOVAL, M.C.** Linear calibration in functional regression models with one of the variances known. 1997. 27p. (RT-MAE-9725)

9726 - **CORDEIRO,G.M.; FERRARI, S.L.P., BOTTER, D.A.; CRIBARI-NETO, F.** Improved maximum likelihood estimation in one-parameter exponential family models. 1997. 16p. (RT-MAE-9726)

9727 - **MACHADO, F.P.** Asymptotic shape for the branching exclusion process. 1997. 16p. (RT-MAE-9727)

9728 - **HO, L.L.; SINGER, J.M.** Generalized least squares methods for bivariate poisson regression. 1997. 12 p. (RT-MAE-9728)

9729 - **MACHADO, F.P.** Large deviations for the number of open clusters per site in long range bond percolation. 1997. 15p. (RT-MAE-9729)

9730 - **BARROSO, L.P., CORDEIRO, G.M., VASCONCELLOS, K.L.P.** Corrected likelihood ratio tests for von mises regression models. 1997. 17p. (RT-MAE-9730)

9731 - **BRESSAUD, X., FERNÁNDEZ, R., GALVES, A.** Speed of $\bar{d}$-convergence for Markov approximations of chains with complete connections. A coupling approach. 1997. 14p. (RT-MAE-9731)

9732 - **FERRARI, S.L.P., SILVA, A.F.** Analytical and resampling-based bias corrections for one-parameter models. 1997. 13P. (RT-MAE-9732)

9733 - **ELIAN, S.N., ANDRÉ, C.D.S., NARULA, S.C.** Influence measure for the minimum sum of absolute errors regression. 1997. 09p. (RT-MAE-9733)

9734 - **PEIXOTO, C.** An Exclusion Process with Metropolis Rate. 1997. 19p. (RT-MAE-9734)

9735 - **ESTEVES, L.G., WECHSLER, S., LEITE, J.G.** DeFinettian Consensus. 1997. 15p. (RT-MAE-9735)

9801 - **GUIOL, H.** Some properties of K-step exclusion processes. 1998. 21p. (RT-MAE-9801)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil