2011

Weighted Scores Method for Regression Models with Dependent Data

Aristidis K. Nikoloulopoulos

Harry Joe

N. Rao Chaganty
Old Dominion University

Original Publication Citation

Nikoloulopoulos, A. K., Joe, H., & Chaganty, N. R. (2011). Weighted scores method for regression models with dependent data. *Biostatistics, 12*(4), 653-665. doi:10.1093/biostatistics/kxr005

Biostatistics (2011), **12**, 4, *pp.* 653–665
doi:10.1093/biostatistics/kxr005
Advance Access publication on March 23, 2011

Weighted scores method for regression models with dependent data

ARISTIDIS K. NIKOLOULOPOULOS*

School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK
a.nikoloulopoulos@uea.ac.uk

HARRY JOE

Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z2

N. RAO CHAGANTY

Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA

SUMMARY

There are copula-based statistical models in the literature for regression with dependent data such as clustered and longitudinal overdispersed counts, for which parameter estimation and inference are straightforward. For situations where the main interest is in the regression and other univariate parameters and not the dependence, we propose a "weighted scores method", which is based on weighting score functions of the univariate margins. The weight matrices are obtained initially fitting a discretized multivariate normal distribution, which admits a wide range of dependence. The general methodology is applied to negative binomial regression models. Asymptotic and small-sample efficiency calculations show that our method is robust and nearly as efficient as maximum likelihood for fully specified copula models. An illustrative example is given to show the use of our weighted scores method to analyze utilization of health care based on family characteristics.

Keywords: Composite likelihood; Copulas; Count data; Estimating equations; Negative binomial.

1. INTRODUCTION

Modeling clustered and longitudinal continuous response data is straightforward based on the multivariate normal (MVN) distribution, applied to a possibly transformed response. This is not the case for discrete responses, such as counts or ordinal variables. Combining copulas (multivariate uniform distributions) with standard univariate models for discrete responses seems to be a promising solution. Exchangeable

*To whom correspondence should be addressed.

copulas could be used to model clustered data. For longitudinal discrete data, copulas with dependence decreasing in time lag can be used.

The MVN copula generated by MVN distribution inherits the useful properties of the latter and thus allows a wide range for dependence and overcomes the drawback of limited dependence of simple parametric families of copulas. Implementation of MVN copula for discrete data (discretized MVN) is possible but not easy because multidimensional integration is needed for computing MVN rectangle probabilities. Therefore, special estimation methods may be required in order to overcome the computational complexities. The use of composite likelihood methods (see Zhao and Joe, 2005; Varin, 2008) can be a potential path toward implementation of discretized MVN and other such models.

In this paper, to generalize existing methodology in the biostatistics literature for regression models with dependent data, we concentrate on inference for the univariate parameters with dependence treated as nuisance, that is, we assume that joint and conditional probabilities are not of interest. The efficiency of estimating the univariate parameters using a composite likelihood based on the sum of univariate log-likelihoods can be low when dependence within clusters is strong. We will improve the efficiency by inserting weight matrices that depend on covariances of the scores assuming a "working model", such as discretized MVN. We call this method "weighted scores".

The weighted scores method is an extension of the generalized estimating equations (GEEs) (Liang and Zeger, 1986) since it can also be applied to families that are not in the class of generalized linear regression models. The GEE method is a nonlikelihood approach based on a "working correlation" matrix. But for nonnormal variables, Pearson's correlations have constraints that depend on the univariate margins, which the GEE method ignores (Sabo and Chaganty, 2010). Furthermore, in general the specified working correlation matrix may not correspond to any multivariate distribution for binary or count data, so regression parameter estimates and p-values may lack a theoretical probabilistic basis. In the absence of a multivariate distribution with specified working correlation matrix, broad claims of consistency of GEE estimates are incorrect (Lee and Nelder, 2009). See Chaganty and Joe (2004, 2006) and Lindsey and Lambert (1998) for further shortcomings of the GEE method. Our new method not only generalizes but also overcomes the theoretical flaws associated with the GEE procedure because our working model is a proper multivariate model and the parameters in the weight matrices are interpretable as latent correlations. Furthermore, we demonstrate that our weighted scores method is highly efficient when compared with the "gold standard" maximum likelihood (ML) estimates arising from proper multivariate models.

To illustrate the method of weighted scores concretely, we use negative binomial regression models to handle cluster and longitudinal count data with overdispersion. For longitudinal count data with overdispersion, Thall and Vail (1990) and Solis-Trapala and Farewell (2005) have proposed covariance models with robust estimation of the covariance matrix, but in these models, negative values for the covariances are not allowed, although multivariate discrete data can exist with some negative associations. Some realistic situations with negative dependence appeared in Chib and Winkelmann (2001). Furthermore, autocorrelation structures such as AR(1), which are very realistic in longitudinal count data, cannot be adapted. Our approach allows flexibility through the MVN copula and more than one form of marginal negative binomial regression as in Cameron and Trivedi (1998) and therefore is superior to the aforementioned methods because it provides (a) robust estimation, (b) more dependence structures, and (c) flexibility in the sense that covariates can be incorporated into both parameters of the negative binomial distribution.

The remainder of the paper proceeds as follows. Section 2 introduces the general theory of weighted scores method, and Section 3 gives details for the special case of negative binomial regression. Section 4 discusses the use of the discretized MVN working model to obtain the weights and asymptotic properties of the estimates. Section 5 studies asymptotic efficiency of our method as compared to ML using other multivariate copula models with negative binomial regression. Section 6 presents an application

of our methodology to analyze dependent overdispersed count data. We conclude this article with some discussion.

2. Weighted scores: general theory

The main idea of weighted scores estimating equations is to write out the score equations for independent data within clusters or panels and then generalize to estimating equations by inserting weight matrices between the matrix of covariates and the vector of scores for regression and nonregression parameters. The general theory is outlined here and made concrete for negative binomial regression models in Section 3.

For ease of exposition, let d be the dimension of a "cluster" or "panel" and n the number of clusters. The theory can be extended to varying cluster sizes. Let p be the number of covariates, that is, the dimension of a covariate vector $\mathbf{x}$. Let $\boldsymbol{\gamma}$ of dimension q be the vector of univariate parameters that are not regression coefficients. The response Y is assumed to have density $f(\cdot;\nu,\boldsymbol{\gamma})$, where $\nu = \eta(\mathbf{x};\boldsymbol{\beta})$ is a function of $\mathbf{x}$ and the p-dimensional regression vector $\boldsymbol{\beta}$. Usually, $\nu = \mathbf{x}^T\boldsymbol{\beta}$ or $\eta(\cdot)$ is a known function.

Suppose that data are $(y_{ij},\mathbf{x}_{ij})$, $j = 1,\ldots,d$, $i = 1,\ldots,n$, where i is an index for individuals or clusters and j is an index for the repeated measurements or within-cluster measurements. The first component of each $\mathbf{x}_{ij}$ is taken as 1 for regression with an intercept. The univariate marginal model for Y_{ij} is $f(\cdot;\nu_{ij},\boldsymbol{\gamma})$, where $\nu_{ij} = \eta(\mathbf{x}_{ij};\boldsymbol{\beta})$. For a multivariate model, one would need a joint distribution of $(Y_{i1},\ldots,Y_{id})$. If for each i, $Y_{i1},\ldots,Y_{id}$ are independent, then the log-likelihood is

$$L = \sum_{i=1}^{n}\sum_{j=1}^{d}\log f(y_{ij};\nu_{ij},\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{d}\ell(\nu_{ij},\boldsymbol{\gamma},y_{ij}),$$

where $\ell(\cdot) = \log f(\cdot)$. The score equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are

$$\begin{pmatrix}\frac{\partial L}{\partial\boldsymbol{\beta}}\\ \frac{\partial L}{\partial\boldsymbol{\gamma}}\end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{d}\begin{pmatrix}\frac{\partial\eta(\mathbf{x}_{ij};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}\begin{pmatrix}\frac{\partial\ell(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\nu_{ij}}\\ \frac{\partial\ell(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\boldsymbol{\gamma}}\end{pmatrix} = \mathbf{0}, \tag{2.1}$$

where $\mathbf{I}_q$ is an identity matrix of dimension q. Let $\mathbf{X}_{ij}^T = \begin{pmatrix}\frac{\partial\eta(\mathbf{x}_{ij};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}$ and $\mathbf{s}_{ij}(\mathbf{a}) = \begin{pmatrix}\frac{\partial\ell(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\nu_{ij}}\\ \frac{\partial\ell(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\boldsymbol{\gamma}}\end{pmatrix}$, where $\mathbf{a}^T = (\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T)$ is the column vector of all $r = p+q$ univariate parameters. The score equations (2.1) can be written as

$$\frac{\partial L}{\partial\mathbf{a}} = \sum_{i=1}^{n}\sum_{j=1}^{d}\mathbf{X}_{ij}^T\mathbf{s}_{ij}(\mathbf{a}) = \sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{s}_i(\mathbf{a}) = \mathbf{0}, \tag{2.2}$$

where $\mathbf{X}_i^T = (\mathbf{X}_{i1}^T,\ldots,\mathbf{X}_{id}^T)$ and $\mathbf{s}_i^T(\mathbf{a}) = (\mathbf{s}_{i1}^T(\mathbf{a}),\ldots,\mathbf{s}_{id}^T(\mathbf{a}))$. The vectors $\mathbf{s}_{ij}(\mathbf{a})$ and $\mathbf{s}_i(\mathbf{a})$ have dimensions $(1+q)$ and $d(1+q)$, respectively. The dimensions of $\mathbf{X}_{ij}$ and $\mathbf{X}_i$ are $(1+q)\times r$ and $d(1+q)\times r$, respectively.

For estimation of $\mathbf{a}$ when $Y_{i1},\ldots,Y_{id}$ are dependent and a multivariate model is not used, we consider weighted scores equations that generalize (2.2) as

$$\mathbf{g} = \mathbf{g}(\mathbf{a}) = \sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{W}_i^{-1}\mathbf{s}_i(\mathbf{a}) = \mathbf{0}, \tag{2.3}$$

where $\mathbf{W}_i$ are invertible $d(1+q) \times d(1+q)$ matrices. In (2.3), for initial analysis, we consider $\mathbf{W}_i$ as fixed. The asymptotic covariance matrix is used to obtain an optimal choice for $\mathbf{W}_i$ for further analysis in the subsequent sections. The asymptotic covariance matrix for the estimator that solves (2.3) is $\mathbf{V} = (-\mathbf{D_g})^{-1}\mathbf{M_g}(-\mathbf{D_g}^T)^{-1}$, where $\mathbf{M_g} = \mathrm{Cov}(\mathbf{g}) = \sum_{i=1}^n \mathbf{X}_i^T\mathbf{W}_i^{-1}\boldsymbol{\Omega}_i[\mathbf{W}_i^T]^{-1}\mathbf{X}_i$ and $-\mathbf{D_g} = E(\partial\mathbf{g}/\partial\mathbf{a}^T) = \sum_{i=1}^n \mathbf{X}_i^T\mathbf{W}_i^{-1}\Psi_i$, for $\mathbf{g}$ in (2.3). Here $\Psi_i = -E(\partial\mathbf{s}_i(\mathbf{a})/\partial\mathbf{a}^T)$ and $\boldsymbol{\Omega}_i = \mathrm{Cov}(\mathbf{s}_i(\mathbf{a}))$. The dimensions of the matrices are as follows: $\mathbf{D_g}$, $\mathbf{M_g}$ are $r \times r$, $\boldsymbol{\Omega}_i$ is $d(1+q) \times d(1+q)$, and Ψ_i is $d(1+q) \times r$. Thus,

$$\mathbf{V} = \left(\sum_{i=1}^n \mathbf{X}_i^T\mathbf{W}_i^{-1}\Psi_i\right)^{-1}\left(\sum_{i=1}^n \mathbf{X}_i^T\mathbf{W}_i^{-1}\boldsymbol{\Omega}_i[\mathbf{W}_i^T]^{-1}\mathbf{X}_i\right)\left(\sum_{i=1}^n \Psi_i^T[\mathbf{W}_i^T]^{-1}\mathbf{X}_i\right)^{-1}.$$

The matrix Cauchy–Schwarz inequality (Chaganty and Joe, 2004) shows that the optimal choice of $\mathbf{W}_i$ satisfies $\mathbf{X}_i^T\mathbf{W}_i^{-1} = \Psi_i^T\boldsymbol{\Omega}_i^{-1}$, leading to

$$\mathbf{V}_{\mathrm{opt}} = \left(\sum_{i=1}^n \Psi_i^T\boldsymbol{\Omega}_i^{-1}\Psi_i\right)^{-1}. \tag{2.4}$$

Thus, the optimal choice of $\mathbf{W}_i$ depends on the "true multivariate model" via Ψ_i and $\boldsymbol{\Omega}_i$.

3. OPTIMAL WEIGHTS FOR NEGATIVE BINOMIAL REGRESSION

For overdispersed count data, the sample variance is larger than the sample mean, that is, the dispersion index (=variance/mean) is greater than 1. The negative binomial distribution $\mathrm{NB}(\tau, \xi)$ allows for overdispersion, and its probability mass function is

$$f(y;\tau,\xi) = \frac{\Gamma(\tau+y)}{\Gamma(\tau)y!}\frac{\xi^y}{(1+\xi)^{\tau+y}}, \quad y = 0, 1, 2, \ldots, \quad \tau > 0,\ \xi > 0,$$

with mean $\mu = \tau\xi$, variance $\tau\xi(1+\xi)$, and dispersion index $1+\xi$. In this parameterization, τ is the convolution parameter and $\xi = 1/\pi - 1$, where π is the Bernoulli parameter.

Cameron and Trivedi (1998) present the $\mathrm{NB}k(\mu,\gamma)$ parameterization, where $\tau = \mu^{2-k}\gamma^{-1}$ and $\xi = \mu^{k-1}\gamma$, $1 \leqslant k \leqslant 2$. For the NB$k$ model, we let $\nu = \log\mu = \mathbf{x}^T\boldsymbol{\beta}$ depend on the covariates $\mathbf{x}$, and consequently either or both the parameters τ and ξ of the negative binomial are covariate dependent. We will primarily use the NB1 parameterization ($\tau = \mu\gamma^{-1}, \xi = \gamma$) and the NB2 parameterization ($\tau = \gamma^{-1}, \xi = \mu\gamma$); the latter is the same as in Lawless (1987). For the NB1 model, τ depends on covariates while $\xi = \gamma$ does not and thus the dispersion index is constant. For the NB2 model, $\tau = 1/\gamma$ is constant while ξ is a function of the covariates and therefore the dispersion index varies with the covariates ranging from $1+\gamma\min[\exp(\mathbf{x}^T\boldsymbol{\beta})]$ to $1+\gamma\max[\exp(\mathbf{x}^T\boldsymbol{\beta})]$. For $1 < k < 2$, one could interpolate between these 2 models using the NBk parameterization, for which the covariates affect both parameters of the negative binomial. Let $\ell(\nu,\gamma,y) = \log f(y;\tau,\xi)$ for the re-parameterized negative binomial model.

Suppose that for each $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant d$, Y_{ij} is distributed as negative binomial with mean μ_{ij} and parameter γ. Assume that $\nu_{ij} = \log(\mu_{ij}) = \mathbf{x}_{ij}^T\boldsymbol{\beta}$. In this case, $\partial\eta(\mathbf{x}_{ij};\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{x}_{ij}$ and $q = 1$. To obtain $\Psi_i = -E(\partial\mathbf{s}_i(\mathbf{a})/\partial\mathbf{a}^T)$, first note that

$$\frac{\partial\mathbf{s}_{ij}(\mathbf{a})}{\partial\mathbf{a}^T} = \begin{pmatrix} \frac{\partial^2\ell}{\partial\nu_{ij}^2}\mathbf{x}_{ij}^T & \frac{\partial^2\ell}{\partial\nu_{ij}\partial\gamma} \\ \frac{\partial^2\ell}{\partial\gamma\,\partial\nu_{ij}}\mathbf{x}_{ij}^T & \frac{\partial^2\ell}{\partial\gamma^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2\ell}{\partial\nu_{ij}^2} & \frac{\partial^2\ell}{\partial\nu_{ij}\partial\gamma} \\ \frac{\partial^2\ell}{\partial\gamma\,\partial\nu_{ij}} & \frac{\partial^2\ell}{\partial\gamma^2} \end{pmatrix}\begin{pmatrix} \mathbf{x}_{ij}^T & 0 \\ \mathbf{0}^T & 1 \end{pmatrix} = \mathbf{D}_{ij}\mathbf{X}_{ij},$$

where $\ell = \ell(\nu_{ij}, \gamma, y_{ij})$. With $\Delta_{ij} = -E(\mathbf{D}_{ij})$, we have

$$\Psi_i = -E\left(\frac{\partial \mathbf{s}_i(\mathbf{a})}{\partial \mathbf{a}^T}\right) = \begin{pmatrix} \Delta_{i1}\mathbf{X}_{i1} \\ \vdots \\ \Delta_{id}\mathbf{X}_{id} \end{pmatrix} = \begin{pmatrix} \Delta_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Delta_{id} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{i1} \\ \vdots \\ \mathbf{X}_{id} \end{pmatrix} = \Delta_i \mathbf{X}_i, \tag{3.1}$$

where $\Delta_i = \mathrm{diag}(\Delta_{i1}, \ldots, \Delta_{id})$ is a symmetric $2d \times 2d$ matrix, as $q = 1$. Explicit expressions for the elements of the matrix Δ_{ij}, the log-likelihood $\ell = \ell(\nu, \gamma, y)$, its derivatives $\partial\ell/\partial\nu$, $\partial\ell/\partial\gamma$, and the negative expectations of the second derivatives $\partial^2\ell/\partial\nu^2$, $\partial^2\ell/\partial\nu\partial\gamma$, $\partial^2\ell/\partial\gamma^2$ for NB1 and NB2 models are given in the supplementary material available at *Biostatistics* online.

Thus, for weighted scores equations of the form (2.3) for dependent observations, the optimal weights $\mathbf{W}_i$ satisfy $\mathbf{X}_i^T\mathbf{W}_i^{-1} = \Psi_i^T\Omega_i^{-1} = \mathbf{X}_i^T\Delta_i^T\Omega_i^{-1}$ or $\mathbf{W}_i = \Omega_i\Delta_i^{-1}$. In the special case of independence within clusters, Δ_i equals $\mathbf{\Omega}_i$ and the optimal $\mathbf{W}_i$ is an identity matrix. In the next section, we propose a preliminary modeling step to get a good estimate for $\mathrm{Cov}(\mathbf{s}_i(\mathbf{a})) = \mathbf{\Omega}_i$ for dependent data.

4. Estimation of weights using a working model

This section presents an estimation method for the weights and asymptotic properties of the weighted scores estimating equations. An approach is to use a working model for the purpose of getting the weight matrices $\mathbf{W}_i$, which might be near optimal for the "true joint distribution".

We select a working model based on the discretized MVN distribution as this allows a wide range of dependence. The discretized MVN model has the following cumulative distribution function (c.d.f.):

$$F(y_1, \ldots, y_d) = \Phi_d(\Phi^{-1}[F_1(y_1)], \ldots, \Phi^{-1}[F_d(y_d)]; \mathbf{R}),$$

where Φ_d denote the standard MVN distribution function with correlation matrix $\mathbf{R}$, Φ is the c.d.f. of the univariate standard normal, and $F_1, \ldots, F_d$ are the univariate distributions. The MVN copula inherits the dependence structure of MVN distribution but lacks a closed-form c.d.f.; this means that likelihood inference might be difficult as d-dimensional integration is required for the multivariate probabilities ($d > 3$) (see Nikoloulopoulos and Karlis, 2009). However, as the optimal weight matrices $\mathbf{W}_i$ depend on covariances of the scores, only the bivariate marginal probabilities of Y_{ij} and Y_{ik}, $j \neq k$, are needed for estimation.

The estimating equations (2.3) based on a "working" discretized MVN take the form

$$\mathbf{g}_1 = \mathbf{g}_1(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_{i,\text{working}}^{-1} \mathbf{s}_i(\mathbf{a}) = \mathbf{0}, \tag{4.1}$$

where $\mathbf{W}_{i,\text{working}}^{-1} = \Delta_i\Omega_{i,\text{working}}^{-1} = \Delta_i(\widetilde{\mathbf{a}})\Omega_i(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}})^{-1}$ is based on the covariance matrix of $\mathbf{s}_i(\mathbf{a})$ computed from the fitted discretized MVN model with estimated parameters $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$. As bivariate normal c.d.f. calculations are needed for the weight matrices (different ones for different clusters), a good approximation that can be quickly computed is important. We used the approximation given by Johnson and Kotz (1972), details of which are described in the supplementary material available at *Biostatistics* online. If the $\mathbf{W}_{i,\text{working}}$ are assumed fixed for the second stage of solving the estimating equations (4.1), then the asymptotic covariance matrix of the solution $\widehat{\mathbf{a}}_1$ is

$$\mathbf{V}_1 = \left(-\mathbf{D}_{\mathbf{g}_1}\right)^{-1} \mathbf{M}_{\mathbf{g}_1} \left(-\mathbf{D}_{\mathbf{g}_1}^T\right)^{-1} \tag{4.2}$$

with

$$-\mathbf{D}_{\mathbf{g}_1} = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_{i,\text{working}}^{-1} \Delta_i \mathbf{X}_i, \quad \mathbf{M}_{\mathbf{g}_1} = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_{i,\text{working}}^{-1} \boldsymbol{\Omega}_{i,\text{true}} (\mathbf{W}_{i,\text{working}}^{-1})^T \mathbf{X}_i,$$

where $\boldsymbol{\Omega}_{i,\text{true}}$ is the "true covariance matrix" of $\mathbf{s}_i(\mathbf{a})$. Alternatively, with $\widetilde{\mathbf{R}}$ from the fitted MVN model assumed fixed, and using the form of the optimal $\mathbf{W}_i$, we consider the following estimating equations in $\mathbf{a}$:

$$\mathbf{g}_2 = \mathbf{g}_2(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^T \Delta_i(\mathbf{a}) \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \mathbf{s}_i(\mathbf{a}) = \mathbf{0}, \tag{4.3}$$

where $\boldsymbol{\Omega}_i(\mathbf{a}, \mathbf{R})$ is the covariance matrix of $\mathbf{s}_i(\mathbf{a})$ based on the discretized MVN model. Solving (4.3) is numerically more intensive than (4.1) because bivariate probabilities $p_{ijk}(y_j, y_k) = \Pr(Y_{ij} = y_j, Y_{ik} = y_k)$ are needed to compute the covariance entries of $\boldsymbol{\Omega}_i(\mathbf{a}, \widetilde{\mathbf{R}})$ for any $\mathbf{a}$ in any iterative numerical method.

The asymptotics for the solution of (4.3) is similar to that of (4.1). Let a_j be a component of $\mathbf{a}$. Then with $\widetilde{\mathbf{R}}$ fixed, we have

$$\begin{aligned}\frac{\partial \mathbf{g}_2}{\partial a_j} &= \sum_{i=1}^{n} \mathbf{X}_i^T \Delta_i(\mathbf{a}) \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \frac{\partial \mathbf{s}_i(\mathbf{a})}{\partial a_j} + \sum_{i=1}^{n} \mathbf{X}_i^T \frac{\partial \Delta_i(\mathbf{a})}{\partial a_j} \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \mathbf{s}_i(\mathbf{a}) \\ &\quad + \sum_{i=1}^{n} \mathbf{X}_i^T \Delta_i(\mathbf{a}) \frac{\partial \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}})}{\partial a_j} \mathbf{s}_i(\mathbf{a}).\end{aligned} \tag{4.4}$$

If the univariate marginal model is correct, then $E[\mathbf{s}_i(\mathbf{a})] = 0$ and only the first term on the right-hand side of (4.4) contributes to $E(\partial \mathbf{g}_2 / \partial \mathbf{a}^T)$. Hence, the asymptotic covariance matrix of the solution $\widehat{\mathbf{a}}_2$ for (4.3) is

$$\mathbf{V}_2 = \left(-\mathbf{D}_{\mathbf{g}_2}\right)^{-1} \mathbf{M}_{\mathbf{g}_2} \left(-\mathbf{D}_{\mathbf{g}_2}^T\right)^{-1} \tag{4.5}$$

with $\mathbf{M}_{\mathbf{g}_2} = \sum_{i=1}^{n} \mathbf{X}_i^T \Delta_i(\mathbf{a}) \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \boldsymbol{\Omega}_{i,\text{true}} \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \Delta_i(\mathbf{a}) \mathbf{X}_i$ and $-\mathbf{D}_{\mathbf{g}_2} = \sum_{i=1}^{n} \mathbf{X}_i^T \Delta_i(\mathbf{a}) \boldsymbol{\Omega}_i^{-1}(\mathbf{a}, \widetilde{\mathbf{R}}) \Delta_i(\mathbf{a}) \mathbf{X}_i$.

To summarize, the steps to obtain parameter estimates and standard errors are as follows. The numerical methods used for various steps are described in the supplementary material available at *Biostatistics* online.

1. Use a discretized MVN model and estimate the parameters using the CL1 composite likelihood method described in Zhao and Joe (2005). Let the parameter estimates be $\widetilde{\mathbf{a}}$ for the univariate parameters and $\widetilde{\mathbf{R}}$ for the correlation matrix.
2. Compute the "working weight matrices" $\mathbf{W}_{i,\text{working}} = \boldsymbol{\Omega}_i(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}}) \Delta_i(\widetilde{\mathbf{a}})^{-1}$, where $\boldsymbol{\Omega}_i(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}})$ is the covariance matrix of univariate scores based on the fitted discretized MVN model.
3. Obtain robust estimates $\widehat{\mathbf{a}}$ of the univariate parameters solving equation (4.1) using working weight matrices $\mathbf{W}_{i,\text{working}}$ or solving (4.3) using $\widetilde{\mathbf{R}}$, with a reliable nonlinear system solver.
4. The robust standard errors for $\widehat{\mathbf{a}}$ are obtained calculating the estimated covariance matrix $\widehat{\mathbf{V}}_1$ of $\widehat{\mathbf{a}}$ based on (4.2) or $\widehat{\mathbf{V}}_2$ based on (4.5) by plugging $\widehat{\mathbf{a}}$ for $\mathbf{a}$ and replacing $\boldsymbol{\Omega}_{i,\text{true}}$ with $\mathbf{s}_i(\widehat{\mathbf{a}}) \mathbf{s}_i^T(\widehat{\mathbf{a}})$. This estimate is similar to the widely used "sandwich" covariance estimator.

5. EFFICIENCY OF WEIGHTED SCORES EQUATIONS

In order to study robustness and efficiency of the weighted scores method, we will use various multivariate copula models as "true" models. A quick overview on copulas, a brief description of some parametric

families of copulas that might be suitable for clustered and longitudinal count data, and their Fisher information matrix are provided in the supplementary material available at *Biostatistics* online.

5.1 *Relative efficiency: comparison based on asymptotic variances*

As a first step to understanding the asymptotic relative efficiency of the weighted scores method for negative binomial regression, we considered copula models with exchangeable and AR(1)-like dependence. For exchangeable dependence structure, the Frank copula in the Archimedean family (Joe, 1997, p 141) with Laplace transform $\phi_{\mathrm{F}}(t) = -\theta^{-1}\log[1-(1-\mathrm{e}^{-\theta})\mathrm{e}^{-t}]$ was used as the "true model". For AR(1)-like dependence, the mixture of max-id copula (Joe and Hu, 1996) with Laplace transform ϕ_{F} and the bivariate Frank copula for the $C_{jk}^{(m)}(\cdot;\theta_{jk})$ was used as the true model. For the marginals, we used both NB1 and NB2 forms of negative binomial regression.

For continuous random variables, dependence as measured by Kendall's tau $\tau = P_c - P_d$, the difference between the probabilities of concordance (P_c) and discordance (P_d), is associated only with the copula parameters. However, for discrete data, the marginal distributions also play a role on dependence and τ does not attain the boundary values of ± 1 because the probability of ties $P_t = 1-(P_c+P_d)$ is positive (see Nikoloulopoulos and Karlis, 2010). Therefore, for discrete data the strength of dependence was measured using the normalized version of Kendall's tau $\tau_{GK} = \frac{P_c-P_d}{P_c+P_d}$ given in Goodman and Kruskal (1954). Various factors related to the dimension, the strength of dependence, and overdispersion were considered in selecting parameters for the models.

A summary of the parameter choices chosen for efficiency calculations is in Table 1. For the covariates and regression parameters, we chose 3 designs: (i) $p = 2, \mathbf{x}_{ij} = (1, x_{1ij})^T$, where x_{1ij} are taken as uniform random variables in the interval $[-1, 1]$, $\beta_0 = -0.5$, and $\beta_1 = 0.5$; (ii) $p = 3, \mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})^T$, where x_{1ij}, x_{2ij} are taken as uniform random variables in the interval $[-1, 1]$, $\beta_0 = -0.5$, and $\beta_1 = \beta_2 = 0.5$; and (iii) $p = 3, \mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})^T$, where x_{1ij} are taken as uniform random variables in the interval $[-1, 1]$ and x_{2ij} are taken as Bernoulli random variables with probability of success $\pi = 0.4$, $\beta_0 = -0.5$, and $\beta_1 = \beta_2 = 0.5$.

For the above copula models and parameter and design selections, we computed the inverse of the Fisher information matrix $\mathcal{I}$, the matrix $\mathbf{V}_{\mathrm{opt}} = \left(\sum_{i=1}^n \Psi_i^T \Omega_i^{-1}\Psi_i\right)^{-1}$, given in (2.4) with Ψ_i as in (3.1), and the matrix $\mathbf{V}_2$ given in (4.5) for both NB1 and NB2 regressions. Note that $\mathcal{I}^{-1}$ is the asymptotic covariance matrix of the ML estimates when the copula model is correctly specified. Since it is common practice to use parametric correlation matrices, we calculated efficiencies for structured latent correlation matrices. For exchangeable dependence, we took $\widetilde{\mathbf{R}}$ as $(1-\rho)\mathbf{I}_d + \rho\mathbf{J}_d$, where $\mathbf{I}_d$ is the identity matrix of order d and $\mathbf{J}_d$ is the $d \times d$ matrix of 1s. For AR(1) dependence, $\widetilde{\mathbf{R}}$ is taken as $(\rho^{|j-k|})_{1\leqslant j,k\leqslant d}$.

Table 1. *Parameter choices in the computation of asymptotic relative efficiencies*

Dependence strength	Exchangeable ($d = 2, 3, 4$)			Longitudinal ($d = 2, 3, 4$)		
	Weak	Moderate	Strong	Weak	Moderate	Strong
τ_{GK}	0.2	0.5	0.8	0.2	0.5	0.8
θ	1.1	3	7	0.1	1.35	4.5
θ_{jk}	—	—	—	1.8	5	20
NB		NB1			NB2	
Overdispersion	Moderate		Large	Moderate		Large
γ	1/2		2	1/2		2

5.2 *Findings on asymptotic relative efficiency*

Representative summaries of findings on the performance of the weighted scores approach are given in Tables 2 and 3 and in the supplementary material available at *Biostatistics* online for three-dimensional ($d = 3$) copula models. We took $n = 500$ to get a good approximation of the asymptotic efficiency. The comparisons are made on the scaled diagonal elements, corresponding to the asymptotic variances of the univariate parameters, of the 3 matrices $\mathcal{I}^{-1}$, $\boldsymbol{V}_{\text{opt}}$, and $\boldsymbol{V}_2$ with different values of ρ.

Table 2 has results for the Frank copula model with moderate dependence; see Table 1 for the values of the dependence parameters. We used design (iii) for the covariates and regression parameters, and the matrix $\widetilde{\mathbf{R}}$ is taken as an exchangeable correlation matrix with parameter ρ varying from 0 to 0.9 in 0.1 increments. Table 3 has results for the mixture of max-id copula model with weak dependence composed by $\phi = \phi_{\text{F}}$ and bivariate Frank copulas with the dependence parameters $\theta_{13} = 0$, $\omega_1 = \omega_3 = -1$, $\omega_2 = 0$ fixed, and θ, θ_{12}, and θ_{23} as in Table 1. We used design (iii), large overdispersion ($\gamma = 2$), and AR(1) structured matrix $\widetilde{\mathbf{R}}$ with parameter ρ ranging from 0 to 0.9 in 0.1 increments. NB1 was used for Table 3 and NB2 was used for Table 2 as the marginal model.

Conclusions from the values in the 2 tables and other computations that we have done are the following.

1. The estimating equations in (2.3) with optimal weight matrices yield estimates that are almost as good as the ML estimates.
2. When the discretized MVN is used as the working model, the weighted scores method yields highly efficient estimates when the parameter ρ is such that the discretized MVN model is quite close to the true model in Kullback–Leibler (K-L) distance.
3. For both negative binomial regressions, the efficiency of the weighted scores method using discretized MVN as a working model is high for the intercept and the dispersion parameter, for a wide

Table 2. *Asymptotic variances, scaled by n, of the marginal parameters for the Frank copula model with moderate dependence and NB2 marginals with moderate dispersion and for the weighted scores (WS) using optimal choice of weights based on the true multivariate model and working weight matrices based on the discretized MVN with exchangeable correlation matrix* $\widetilde{\mathbf{R}}$ *with parameter* ρ. *Efficiencies with respect to ML are shown in parentheses*

Method (covariance)	$n\text{Var}(\beta_0)$	$n\text{Var}(\beta_1)$	$n\text{Var}(\gamma)$
ML ($\mathcal{I}^{-1}$)	1.178 (1.00)	1.804 (1.00)	4.234 (1.00)
Optimal WS ($\boldsymbol{V}_{\text{opt}}$)	1.206 (0.98)	1.846 (0.98)	4.339 (0.98)
WS with discretized MVN ($\boldsymbol{V}_2(\rho)$)			
$\rho = 0.0$	1.210 (0.97)	2.213 (0.82)	4.346 (0.97)
$\rho = 0.1$	1.208 (0.97)	2.043 (0.88)	4.345 (0.98)
$\rho = 0.2$	1.208 (0.98)	1.938 (0.93)	4.343 (0.98)
$\rho = 0.3$	1.207 (0.98)	1.883 (0.96)	4.341 (0.98)
$\rho = 0.4^{\dagger}$	1.208 (0.98)	1.866 (0.97)	4.340 (0.98)
$\rho = 0.5$	1.208 (0.98)	1.881 (0.96)	4.344 (0.96)
$\rho = 0.6$	1.210 (0.97)	1.923 (0.94)	4.368 (0.97)
$\rho = 0.7$	1.212 (0.97)	1.989 (0.91)	4.446 (0.95)
$\rho = 0.8$	1.218 (0.97)	2.088 (0.86)	4.704 (0.90)
$\rho = 0.9$	1.235 (0.95)	2.259 (0.80)	5.680 (0.75)

†For this ρ value, the discretized MVN working model is close to the true model in K-L distance.

Table 3. *Asymptotic variances, scaled by n, of the marginal parameters for the trivariate mixture of max-id copula model composed by ϕ_F with weak dependence and NB1 marginals with large dispersion and for the weighted scores (WS) using optimal choice of weights based on the true multivariate model and working weight matrices based on the discretized MVN with AR(1) correlation matrix $\widetilde{\mathbf{R}}$ with parameter ρ. Efficiencies with respect to ML are shown in parentheses*

Method (covariance)	$n\mathrm{Var}(\beta_0)$	$n\mathrm{Var}(\beta_1)$	$n\mathrm{Var}(\beta_2)$	$n\mathrm{Var}(\gamma)$
ML ($\mathcal{I}^{-1}$)	2.481 (1.00)	2.796 (1.00)	3.520 (1.00)	17.827 (1.00)
Optimal WS ($\boldsymbol{V}_{\mathrm{opt}}$)	2.467 (1.00)	2.799 (1.00)	3.526 (1.00)	17.772 (1.00)
WS with discretized MVN ($\boldsymbol{V}_2(\rho)$)				
$\rho = 0.0$	2.487 (1.00)	2.856 (0.98)	3.602 (0.98)	17.777 (1.00)
$\rho = 0.1$	2.472 (1.00)	2.811 (0.99)	3.543 (0.99)	17.777 (1.00)
$\rho = 0.2^{\dagger}$	2.469 (1.00)	2.803 (1.00)	3.531 (1.00)	17.781 (1.00)
$\rho = 0.3$	2.481 (1.00)	2.834 (0.99)	3.574 (0.99)	17.795 (1.00)
$\rho = 0.4$	2.509 (0.99)	2.908 (0.96)	3.675 (0.96)	17.833 (1.00)
$\rho = 0.5$	2.559 (0.97)	3.028 (0.92)	3.841 (0.92)	17.925 (0.99)
$\rho = 0.6$	2.635 (0.94)	3.201 (0.87)	4.079 (0.86)	18.128 (0.98)
$\rho = 0.7$	2.745 (0.90)	3.436 (0.81)	4.403 (0.80)	18.562 (0.96)
$\rho = 0.8$	2.914 (0.85)	3.759 (0.74)	4.844 (0.73)	19.517 (0.91)
$\rho = 0.9$	3.276 (0.76)	4.245 (0.66)	5.490 (0.64)	22.050 (0.81)

†For this ρ value, the discretized MVN working model is close to the true model in K-L distance.

range of ρ values. However, the efficiencies are not as high for the remaining regression coefficients. This indicates that incorrect choice of ρ could lead to significant loss of efficiency.

4. Efficiencies are quite flat in an interval of ρ that depends on the strength of dependence within the clusters. This provides partial justification for using (4.2) for the estimated covariance matrix of the weighted scores estimator.

5.3 *Small-sample efficiency based on simulation studies*

To gauge the small-sample efficiency of the weighted scores method, we performed several simulation studies using the copula models with parameter choices and design matrices as in Table 1 and Section 5.1. We report here typical results from these experiments. We randomly generated $B = 10^4$ samples of size $n = 500, 300, 100$ from the trivariate exchangeable Frank copula with moderate dependence and NB2 regression with moderate dispersion and design (i) as in Section 5.1. Table 4 contains the parameter values, the bias, variance (Var), and mean square errors of the ML estimates and weighted scores, along with the average of their theoretical variances. The theoretical variance of the ML estimate is obtained via the gradients and the Hessian computed numerically during the maximization process. The weighted scores estimates were obtained assuming that the $\boldsymbol{W}_{i,\mathrm{working}}$ are fixed for the second stage of solving the estimating equations in (4.3). The reported theoretical variances for the weighted scores method are from $\boldsymbol{V}_1$ as in (4.2). The variances from $\boldsymbol{V}_2$ were similar. Therefore, it is adequate to use $\boldsymbol{V}_1$ for estimating the standard errors of the weighted scores estimates.

It is clear from Tables 2 and 4 ($n = 500$) that variances computed from the simulations are similar to the asymptotic variances for both the ML and the weighted scores method. For example, for $\rho = 0.4$ in Table 2, $n\mathrm{Var}(\beta_0)$ is 1.208 which is approximately equal to 1.217 in Table 4. Note that for $\rho = 0.4$, the discretized MVN working model is close to the Frank copula model under moderate dependence with

Table 4. *Small sample of sizes $n = 500, 300, 100$ simulations (10^4 replications) and resulted biases and mean square errors (MSE) and variances (Var), along with average theoretical variances scaled by n, for the ML of the marginal parameters for the trivariate Frank copula model with moderate dependence and NB2 marginals with moderate dispersion and for the weighted scores based on the discretized MVN with exchangeable correlation matrix* $\mathbf{R}$ *with parameter* ρ *as estimated by the CL1 method*

		ML				Weighted scores			
	n	nBias	nVar	nMSE	$n\bar{\boldsymbol{V}}(\cdot)$	nBias	nVar	nMSE	$n\bar{\boldsymbol{V}}_1(\cdot)$
$\beta_0 = -0.5$	500	−0.82	1.19	1.19	1.17	−0.83	1.22	1.22	1.20
	300	−1.11	1.18	1.19	1.17	−1.10	1.22	1.22	1.20
	100	−1.03	1.22	1.23	1.18	−1.01	1.24	1.26	1.20
$\beta_1 = 0.5$	500	0.62	1.81	1.81	1.80	0.61	1.86	1.86	1.85
	300	0.29	1.79	1.79	1.80	0.24	1.85	1.85	1.85
	100	0.20	1.90	1.90	1.84	0.19	1.95	1.95	1.85
$\gamma = 0.5$	500	−0.99	4.36	4.36	4.36	−0.74	4.46	4.46	4.42
	300	−0.94	4.44	4.44	4.40	−0.71	4.55	4.55	4.42
	100	−0.62	4.48	4.48	4.58	−0.42	4.61	4.61	4.43

respect to the K-L distance, a measure that is useful to make theoretical likelihood comparisons between 2 models.

For the models of Section 5.1, the K-L distance slowly increases as the NB mean and the dependence increase and it is less than 0.1 if the NB mean is less than 4 and the dependence is less than 0.8 as measured with τ_{GK}. Our calculations show similar magnitude of K-L distance of copula models with discretized MVN for exchangeable, AR(1), and unstructured dependence. Therefore, we expect the patterns seen for exchangeable and AR(1) dependence to hold for unstructured dependence. In conclusion, the simulated variances for samples of size $n = 500, 300, 100$ show that the weighted scores method is almost as good as ML.

6. ANALYSIS OF UTILIZATION OF HEALTH CARE COUNT DATA

This section illustrates the application of the weighted scores method to Riphahn, Wambach, and Million (RWM) data (Riphahn *and others*, 2003), consisting of 7293 families from former West Germany and of German nationality observed from 1 to 7 times during the years 1984–1988, 1991, and 1994. The data are available for download at http://econ.queensu.ca/jae/2003-v18.4/. The focus of the original survey was to study the role of public, private, and add-on health insurance on the intensity and utilization of health care facilities. The use of health care facilities was measured by 2 primary count response variables, the number of visits to a doctor (DocVis) within the last quarter prior to the survey and the number of inpatient hospital visits (HospVis) within a given calendar year. Besides these variables the survey has a number of explanatory variables. A complete list of variables and descriptive statistics can be found in Table 1 in Greene (2008).

Our primary goal is not a complete analysis of the RWM data but to show the use of the weighted scores in regressing the count variable (DocVis) panel on the explanatory variables. For our analysis, we have selected a further subset of the RWM data consisting of 1154 families which had complete data for the 5 years 1984–1988 and then to emphasize that our method does not depend on a constant cluster size d, some observations were dropped randomly for the dependent variable. Thus, the data are unbalanced and the missing values can be assumed to be missing completely at random.

Table 5. *Weighted scores (WS) estimates and standard errors (SEs) for the utilization of health care count data*

CL1 estimates of the dependence parameters										
Marginal	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{15}	ρ_{23}	ρ_{24}	ρ_{25}	ρ_{34}	ρ_{35}	ρ_{45}
NB1	0.46	0.42	0.29	0.24	0.31	0.32	0.36	0.29	0.21	0.28
NB2	0.46	0.41	0.32	0.25	0.31	0.35	0.37	0.31	0.23	0.28

WS estimates of the univariate parameters								
	NB1				NB2			
Covariates†	Estimate	Robust SE	Z	Pr > $\lvert Z \rvert$	Estimate	Robust SE	Z	Pr > $\lvert Z \rvert$
Intercept	0.343	0.294	1.17	0.24	0.635	0.367	1.76	0.08
Sex	0.373	0.091	4.10	0.00	0.248	0.107	2.28	0.02
Age	0.018	0.004	4.14	0.00	0.018	0.005	3.55	0.00
Hsat	−0.115	0.014	−8.33	0.00	−0.147	0.019	−7.60	0.00
Handper	0.006	0.001	5.25	0.00	0.006	0.002	3.22	0.00
Univ	−0.439	0.305	−1.44	0.15	−0.707	0.306	−2.33	0.02
Public	0.246	0.191	1.29	0.20	0.213	0.247	0.85	0.39
Addon	−0.089	0.247	−0.36	0.72	−0.371	0.228	−1.63	0.10
γ	3.910	0.332	11.76	0.00	1.263	0.097	13.02	0.00

†Sex = indicator of female, that is, 1 for female, 0 for male; age = age in years; hsat = health satisfaction, on a scale of 0–10; handper = degree of handicap in percent, 0–100; univ = binary indicator that highest schooling degree is university; public = indicator of public health insurance; addon = indicator of add-on insurance.

Table 5 gives the estimates of the univariate parameters, along with the dependence parameters obtained using the weighted scores method. As there is no reason to assume a structured correlation between the 5 (1984–1988) repeated measurements, in the first step of the method the unstructured correlation matrix $\mathbf{R}$ in the MVN copula is estimated using the CL1 method in Zhao and Joe (2005). These CL1 estimates are listed in the first rows of the table. Next, we used this estimate of $\widetilde{\mathbf{R}}$ and solved the weighted scores equations (4.1) to obtain estimates $\widehat{\mathbf{a}}_1$ of the univariate parameters for both NB1 and NB2 regressions. The parameter estimates along with robust standard errors computed using $\widehat{\boldsymbol{V}}_1$ are also presented in Table 5. Since the parameters are the same for the 2 models, we can use the bivariate log-likelihood at CL1 estimates as a rough diagnostic measure for goodness of fit between the 2 NB models. This quantity was −7835.9 for NB2 and −7753.4 for NB1, and thus NB1 seems to be a better fit for the data.

Interestingly, the coefficient of hsat is negative and highly significant, indicating that families with high health satisfaction rating seem to be making less frequent trips to the doctors office. Further public and add-on health insurance has insignificant effect on the doctor visits as expected since families are more likely to make doctor visits based on their health care needs as opposed to the type of insurance they carry.

7. Discussion

In this paper, we have studied weighted scores as an estimating equations approach based on weighting the scores of the marginal distributions to account for the dependence in repeated or clustered measurements. The weighted scores method with fixed weight matrices leads to unbiased estimating equations if the univariate model is correct, and the efficiency depends on the choice of weight matrices. On the other hand, the ML equations also lead to unbiased estimating equations if the univariate model and the dependence are modeled correctly; however, the ML estimates could be biased if the univariate model is correct but

dependence is modeled incorrectly. Hence, the weighted scores method is robust to dependence if the main interest is the univariate regression parameters.

Some multivariate models are nearly indistinguishable from each other based on the moderate level of dependence in response variables and sample sizes usually seen for real data. If there is strong dependence, different copula models can be more easily discriminated and there might be a better working model than discretized MVN. In terms of the Akaike information criteria, our empirical experience is that the MVN copula model with discrete margins provides the best or nearly the best fit, so the working model based on the discretized MVN distribution to compute weight matrices can satisfactorily account for the dependence in the data.

ACKNOWLEDGMENTS

Thanks to the associate editor and 2 referees for comments leading to an improved presentation. *Conflict of Interest:* None declared.

FUNDING

Natural Sciences and Engineering Research Council Canada, Discovery Grant 8698.

REFERENCES

CAMERON, A. C. AND TRIVEDI, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

CHAGANTY, N. R. AND JOE, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B* **66**, 851–860.

CHAGANTY, N. R. AND JOE, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93**, 197–206.

CHIB, S. AND WINKELMANN, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics* **19**, 428–435.

GOODMAN, L. A. AND KRUSKAL, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732–764.

GREENE, W. (2008). Functional forms for the negative binomial model for count data. *Economic Letters* **99**, 585–590.

JOE, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

JOE, H. AND HU, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis* **57**, 240–265.

JOHNSON, N. L. AND KOTZ, S. (1972). *Continuous Multivariate Distributions*. New York: Wiley.

LAWLESS, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**, 209–225.

LEE, Y. AND NELDER, J. A. (2009). Likelihood inference for models with unobservables: another view. *Statistical Science* **24**, 255–269.

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

LINDSEY, J. K. AND LAMBERT, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* **17**, 447–469.

NIKOLOULOPOULOS, A. K. AND KARLIS, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference* **139**, 3878–3890.

NIKOLOULOPOULOS, A. K. AND KARLIS, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics* **37**, 1555–1568.

RIPHAHN, R. T., WAMBACH, A. AND MILLION, A. (2003). Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of Applied Econometrics* **18**, 387–405.

SABO, R. T. AND CHAGANTY, N. R. (2010). What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. *Statistics in Medicine* **29**, 2501–2507.

SOLIS-TRAPALA, I. L. AND FAREWELL, V. T. (2005). Regression analysis of overdispersed correlated count data with subject specific covariates. *Statistics in Medicine* **24**, 2557–2575.

THALL, P. F. AND VAIL, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.

VARIN, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* **92**, 1–28.

ZHAO, Y. AND JOE, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* **33**, 335–356.

[*Received May 10, 2010; revised January 23, 2011; accepted for publication January 24, 2011*]