Published in final edited form as:

Stat Methods Med Res. 2020 August ; 29(8): 2087–2099. doi:10.1177/0962280219884980.

Modification of the generalized quasi-likelihood model in the analysis of the Add Health study

Katherine E. Irimata[1,*], **Jeffrey R. Wilson**[2]

[1]Division of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA

[2]Department of Economics, Arizona State University, Tempe, AZ, USA

Abstract

The relationship between the mean and variance is an implicit assumption of parametric modeling. While many distributions in the exponential family have a theoretical mean-variance relationship, it is often the case that the data under investigation are correlated, thus varying from the relation. We present a generalized method of moments estimation technique for modeling certain correlated data by adjusting the mean-variance relationship parameters based on a canonical parameterization. The proposed mean-variance form describes overdispersion using two parameters and implements an adjusted canonical parameter which makes this approach feasible for all distributions in the exponential family. Test statistics and confidence intervals are used to measure the deviations from the mean-variance relation parameters. We use the modified relation as a means of fitting generalized quasi-likelihood models to correlated data. The performance of the proposed modified generalized quasi-likelihood model is demonstrated through a simulation study and we highlight the importance of accounting for overdispersion in the evaluation of adolescent obesity data collected from a U.S. longitudinal study.

Keywords

Canonical parameter; correlation; generalized linear models; generalized method of moments; overdispersion

1. Introduction

As a common statistical measure, the variance is often relied on to evaluate the model fit and to understand the differences between the responses through the construction of test statistics and confidence intervals. The form of the variance is often assumed based on the underlying distribution of the responses. In fact, the variance is related to the mean for most

[*]Corresponding author: Katherine Irimata, Division of Research and Methodology, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD, USA, kirimata@cdc.gov.

Publisher's Disclaimer: Disclaimer: The work for this paper was conducted while the first author was at Arizona State University. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Declaration of Conflicting Interests
The Authors declare that there is no conflict of interest.

distributions in the exponential family. However, while the responses may be on a certain scale or resemble a certain distribution, extraneous variation can impact the mean-variance relationship. Extraneous variation, or so-called overdispersion, is often present in longitudinal or clustered data arising from a hierarchical data structure.

Ignoring overdispersion in the fit of correlated data results in summary statistics, including test statistics, with a larger variance than expected.[1] It often leads to a loss of efficiency in using statistics appropriate for the single-parameter family.[2] Studies have shown that ignoring overdispersion and thereby misspecifying the model biases the covariate effects and greatly impacts the standard error of the coefficients.[3, 4] While underdispersion, the case when the variation is smaller than expected, may occur and also impacts the accuracy of the analysis when it is not appropriately specified, McCullagh and Nelder[5] have suggested that overdispersion may be the norm. Various methods have been proposed to identify the underlying variation and provide corrections to improve estimates of the variance.[6, 7]

Overdispersion or underdispersion is often identified by estimating the parameters in the mean-variance relationship and measuring the deviations from the theoretical values under the assumed distribution. Kukush et al.[8] considered a pair of mean and variance functions with a common parameter vector θ estimated using an extended quasi-score function. Tsou[9] considered two parameters (ψ, λ) in a parametric robust method of determining the mean-variance relationship through estimation of the power λ with an adjusted robust log likelihood method for fixed values of ψ. In addition, researchers have developed methods to test for overdispersion in proportions[10] and score test statistics for overdispersed Poisson and binomial models.[11] Xiang et al.[12] provided a score test for overdispersion in a zero-inflated Poisson mixed regression model. Yang, Hardin, and Addy[13] modified the score statistic to test overdispersion in the zero–inflated generalized Poisson mixed model. While these tests work well for identifying overdispersion, current parameterizations are limited to one parameter or are only applicable to distributions that have a particular form for the variance.

Overdispersed data are analyzed with appropriate statistical models such as generalized estimating equations, generalized linear mixed models, and joint modeling of the mean and dispersion.[14] Generalized estimating equations account for correlation through the selection of a covariance structure for the correlated responses.[15] Generalized linear mixed models have been used to model overdispersion in non-normal data.[16] These models incorporate random effects, through random intercepts and random slopes, to account for correlation due to clustering.[17] The joint modeling of the mean and the variance uses an additional dispersion submodel to address the overdispersion in a generalized linear model context.[18] Joint modeling allows one to simultaneously model both the mean and the variance through submodels. This technique has been extended to consider joint modeling in hierarchical generalized linear model structures.[19, 20]

Quasi-likelihood models are useful in cases where the underlying distribution is unspecified.[21] This modeling technique relaxes the distributional assumption in the random component and instead relies on the specification of a mean-variance function. The regression parameter estimates and standard errors are obtained from the specified mean-variance relationship and estimates of the covariance matrix in a quadratic form. The quasi-likelihood approach possesses many good properties, including unbiased estimates and small standard errors as compared to alternative methods.[22] While the quasi-likelihood method is appropriate for evaluating overdispersed data, the form of the variance has been limited to a single multiplicative overdispersion parameter.

This paper proposes a modified generalized quasi-likelihood (MGQL) model which utilizes a canonical two-parameter mean-variance relation. The proposed canonical parameterization is flexible and can be used to represent the form of the variance for any distribution in the exponential family. The incorporation of this mean-variance relationship in the MGQL extends quasi-likelihood models to describe a larger class of variance functions in the analysis of correlated data.

In Section 2, we review mean-variance relationships in the exponential family and the generalized quasi-likelihood approach to estimate the regression parameters and variance components in the analysis of clustered data.[23] In Section 3, we present the canonical parameterization of the mean-variance relationship. We provide a generalized method of moments (GMM) approach to estimate the mean-variance parameters and introduce a test to identify overdispersion or underdispersion. We propose a model that incorporates the mean-variance relationship in the form of a modified generalized quasi-likelihood. In Section 4, a simulation study is utilized to demonstrate the performance of the MGQL model. In Section 5, the MGQL model is used to analyze data collected through the National Longitudinal Study of Adolescent to Adult Health (Add Health).[24] This study collected health information on adolescents over four waves of interviews, and is highly correlated due to the nested structure of the longitudinal study. We demonstrate the use of the MGQL to appropriately account for overdispersion in the evaluation of risk factors associated with obesity.

2. Background

2.1 Mean-Variance Relation Parameters

Let the vector of observations $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ be realizations of a set of independent random variables $\boldsymbol{Y}$ with means $\boldsymbol{\mu}$ related to a set of k covariates $\boldsymbol{X} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_k^T\right)$ through a function $g(\cdot)$ such that $E(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ and $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{X}\beta$. The estimates of the regression parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^T$ can be obtained using estimating equations from the likelihood. Let the joint probability function of Y_i with known functions a, b, and c and known dispersion parameter $\boldsymbol{\phi}$[5] be

$$f(y; \theta, \phi) = exp\left[\frac{(y\theta - b(\theta))}{a(\phi)} + c(y, \phi)\right].$$

If θ is unknown, we have a two-dimensional exponential family with log likelihood

$$l(\theta, \phi \mid y) = \frac{(y\theta - b(\theta))}{a(\phi)} + c(y, \phi).$$

Using the expectation of the derivative of the likelihood

$$E\left(\frac{\partial l(\theta,\phi|y)}{\partial\theta}\right)=0$$

and the property

$$E\left(\frac{\partial^2 l(\theta,\phi|y)}{\partial\theta^2}\right)+E\left(\frac{l(\theta,\phi|y)}{\partial\theta}\right)^2=0$$

yields the expected value $E(\boldsymbol{Y})=b'(\theta)$ and the variance $var(\boldsymbol{Y})=b''(\theta)a(\phi)$ where $b'(\cdot)$ denotes the first derivative and $b''(\cdot)$ denotes the second derivative. Thus, both b' and b'' are functions of the canonical parameter θ. The mean and the variance are related through the first derivative and second derivative of the function $b(\theta)$. The variance of the observations is a product of a function of the canonical parameter θ and a function of the dispersion parameter ϕ.

The Poisson distribution and the binomial distribution are members of the exponential family and are commonly used to analyze count data and binary data, respectively. The Poisson distribution has probability mass function

$$f(y;\alpha)=exp(y\,log\,\alpha-\alpha-logy!)$$

with $a(\phi)=1$, $b(\theta)=\alpha=\exp(\theta)$, $c(y,\phi)=-\log y!$, and canonical parameter $\theta=\log(\alpha)$. Thus, the expected mean and variance under the Poisson distribution are equal to α. The binomial distribution has probability distribution function

$$f(y;m,p)=\binom{m}{y}p^y(1-p)^{m-y},$$

with $a(\phi)=1$, $b(\theta)=m\log(1+\exp(\theta))$, $c(y,\phi)=\log\binom{m}{y}$ and the canonical parameter $\theta=\log\left(\frac{p}{1-p}\right)$. Under the binomial distribution, the expected mean is mp and the expected variance is $mp(1-p)$.[5, 25]

2.2 Generalized Quasi-likelihood Models

Generalized quasi-likelihood models use the specification of the mean-variance relationship to evaluate correlated data. Consider vectors of correlated observations $\boldsymbol{y}_1,\dots\boldsymbol{y_n}$ for $i=1,\dots n$ where $\boldsymbol{y_i}=\left(y_{i1},\dots.y_{in_i}\right)$. The correlated observations y_{ij} for $j=1,\dots,n_i$ where n_i is the sample size of the ith vector comes from a distribution in the exponential family with link function $g(\cdot)$ such that

$$g\left(\mu_{ij}\right)=\eta_{ij}=\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\sigma\xi_i$$

for $\xi_i \sim N(0, 1)$ as $\xi_i = \alpha_i/\sigma$ for k covariates and random effects $\alpha_i \sim N\left(0, \sigma^2\right)$ for cluster i. We estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ using GQL. Let the response vector be

$$\boldsymbol{S_i} = \left(y_i^{\mathrm{T}}, u_i^{\mathrm{T}}\right)$$

for cluster i, where $\boldsymbol{y_i}^{\mathrm{T}} = \left(y_{i1}, \ldots, y_{in_i}\right)$ and $\boldsymbol{u_i}^{\mathrm{T}} = \left(\boldsymbol{u}_{\boldsymbol{i}1}^{\mathrm{T}}, \boldsymbol{u}_{\boldsymbol{i}2}^{\mathrm{T}}\right)$ contains the pairwise products for $\boldsymbol{u}_{\boldsymbol{i}}1 = \left(y_{i1}^2, \ldots, y_{in_i}^2\right)$ and $\boldsymbol{u}_{\boldsymbol{i}}2 = \left(y_{i1}y_{i2}, \ldots, y_{ij}y_{ij*}, \ldots, y_{i(n_i-1)}y_{in_i}\right)$. Let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \sigma\right)^T$ and $\boldsymbol{M_i}(\theta)$ be the mean of the response vector $\boldsymbol{S_i}$. Let $\boldsymbol{\Omega_i}(\theta)$ be the covariance matrix for $\boldsymbol{S_i}$ with elements ω_{ij}. Then, the set of generalized quasi-likelihood estimating equations,

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{M}_{\boldsymbol{i}}'(\theta)}{\partial \theta} \boldsymbol{\Omega}_{\boldsymbol{i}}^{-1}(\theta)\left[\boldsymbol{S_i} - \boldsymbol{M_i}(\theta)\right] = 0 \tag{2.1}$$

provides GQL estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$.[21, 23] The mean of the response vector, $\boldsymbol{M_i}(\theta)$, is evaluated as

$$\boldsymbol{M_i}(\theta) = E(\boldsymbol{S_i}) = E\left(Y_{i1}, \ldots, Y_{in_i}, Y_{i1}^2, \ldots, Y_{in_i}^2, Y_{i1}Y_{i2}, \ldots, Y_{i(n_i-1)}Y_{n_i}\right)$$

and

$$E\left(Y_{ij}\right) = \mu_{ij}(\theta) = E\left[g_{(1)}\left(x_{ij}^T\beta + \sigma\xi\right)\right]$$

$$E\left(Y_{ij}^2\right) = m_{ijj}(\theta) = E\left[g_{(2)}\left(x_{ij}^T\beta + \sigma\xi\right)\right]$$

$$E\left(Y_{ij}Y_{ik}\right) = m_{ijk}(\theta) = E\left[g_{(1)}\left(x_{ij}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{ik}^T\beta + \sigma\xi\right)\right]$$

where the functions $g_{(r)}\left(\eta_{ij}\right)$ are the r^{th} finite moments of y_{ij}. The partial derivative matrix $\frac{\partial \boldsymbol{M}_{\boldsymbol{i}}'(\theta)}{\partial \theta}$ has dimension $(p+1) \times \{n_i(n_i+1)/2\}$, with partial derivatives

$$\frac{\partial \mu_{ij}(\theta)}{\partial \beta} = E\left[\tilde{g}_{(1)}\left(x_{ij}^T\beta + \sigma\xi\right)\right]x_{ij}^T$$

$$\frac{\partial m_{ijj}(\theta)}{\partial \beta} = E\left[\tilde{g}_{(2)}\left(x_{ij}^T\beta + \sigma\xi\right)\right]x_{ij}^T$$

$$\frac{\partial m_{ijk}(\theta)}{\partial \beta} = E\left[\tilde{g}_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)g_{(1)}\left(x_{ik}^{T}\beta + \sigma\xi\right)\right]x_{ij}^{T} + E\left[g_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)\tilde{g}_{(1)}\left(x_{ik}^{T}\beta\right.\right.$$
$$\left.\left. + \sigma\xi\right)\right]x_{ik}^{T}$$

$$\frac{\partial \mu_{ij}(\theta)}{\partial \sigma} = E\left[\xi\tilde{g}_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)\right]x_{ij}^{T}$$

$$\frac{\partial m_{ijj}(\theta)}{\partial \sigma} = E\left[\xi\tilde{g}_{(2)}\left(x_{ij}^{T}\beta + \sigma\xi\right)\right]x_{ij}^{T}$$

$$\frac{\partial m_{ijk}(\theta)}{\partial \sigma} = E\left[\xi\left\{\tilde{g}_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)g_{(1)}\left(x_{ik}^{T}\beta + \sigma\xi\right) + g_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)\tilde{g}_{(1)}\left(x_{ik}^{T}\beta + \sigma\xi\right)\right\}\right]$$

where $\tilde{g}_{(r)}(\cdot)$ as the first derivative of $g_{(r)}(\cdot)$. Then the covariance matrix is

$$\boldsymbol{\Omega_i} = \begin{bmatrix} \boldsymbol{\Sigma_i} & \boldsymbol{P_i} \\ \boldsymbol{P_i'} & \boldsymbol{Q_i} \end{bmatrix}$$

where $\Sigma_{\mathbf{i}} = cov(Y_i)$, $\mathbf{P_i} = cov\left(Y_i, U_i^{\mathrm{T}}\right)$ and $\mathbf{Q_i} = cov(U_i)$. The diagonal elements of $\boldsymbol{\Sigma}_{\mathbf{i}}$ are the variances of Y_i such that $\sigma_{ijj} = Var\left(Y_{ij}\right) = m_{ijj}(\theta) - \mu_{ij}^{2}(\theta)$ with off diagonal elements $\sigma_{ijk} = Cov\left(Y_{ij}, Y_{ik}\right) = m_{ijk}(\theta) - \mu_{ij}(\theta)\mu_{ik}(\theta)$. The matrix $\mathbf{P_i}$ is of dimension $n_i \times \{n_i(n_i + 2)/2\}$ and contains $cov\left(Y_{ij}, Y_{ij}^{2}\right)$, $cov\left(Y_{ij}, Y_{ij}Y_{il}\right)$ and $cov\left(Y_{ij}, Y_{ik}Y_{il}\right)$. For $j = k = l$,

$$cov\left(Y_{ij}, Y_{ij}^{2}\right) = p_{ijjj}(\theta) = E\left[g_{(3)}\left(x_{ij}^{T}\beta + \sigma\xi\right)\right] - \mu_{ij}(\theta)m_{ijj}(\theta).$$

For $j = k \neq l$ and $j = l \neq k$, the covariance elements are

$$cov\left(Y_{ij}, Y_{ij}Y_{il}\right) = E\left(Y_{ij}^{2}Y_{il}\right) - \mu_{ij}(\theta)m_{ijl}(\theta)$$

$$= E\left[g_{(2)}\left(x_{ij}^{T}\beta + \sigma\xi\right)g_{(1)}\left(x_{il}^{T}\beta + \sigma\xi\right)\right] - \mu_{ij}(\theta)m_{ijl}(\theta).$$

For $j \neq k \neq l$,

$$cov\left(Y_{ij}, Y_{ik}Y_{il}\right) = p_{ijkl}(\theta)$$

$$= E\left[g_{(1)}\left(x_{ij}^{T}\beta + \sigma\xi\right)g_{(1)}\left(x_{ik}^{T}\beta + \sigma\xi\right)g_{(1)}\left(x_{il}^{T}\beta + \sigma\xi\right)\right] - \mu_{ij}(\theta)m_{ikl}(\theta).$$

In the covariance matrix, $\boldsymbol{Q_i}$ contains $cov\left(Y_{ij}Y_{ik}, Y_{il}Y_{iw}\right)$ with dimension $\{n_i(n_i+1)/2\} \times \{n_i(n_i+1)/2\}$. For $j = k = l = m$,

$$cov\left(Y_{ij}^2, Y_{ij}^2\right) = q_{ijjjj}(\theta) = E\left[g_{(4)}\left(x_{ij}^T\beta + \sigma\xi\right)\right] - m_{ijj}^2(\theta)\,.$$

For $j = k \neq l \neq w$,

$$cov\left(Y_{ij}^2, Y_{il}Y_{iw}\right) = q_{ijjlw}(\theta)$$

$$= E\left[g_{(2)}\left(x_{ij}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{il}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{iw}^T\beta + \sigma\xi\right)\right] - m_{ijj}(\theta)m_{ilw}(\theta)\,.$$

For $j \neq k \neq l \neq w$,

$$cov\left(Y_{ij}Y_{ik}, Y_{il}Y_{iw}\right) = q_{ijklw}$$

$$= E\left[g_{(1)}\left(x_{ij}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{ik}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{il}^T\beta + \sigma\xi\right)g_{(1)}\left(x_{iw}^T\beta + \sigma\xi\right)\right] - m_{ijk}(\theta)m_{ilw}(\theta)\,.$$

The quasi-likelihood estimate $\hat{\boldsymbol{\theta}}_{QL} = \left(\hat{\boldsymbol{\beta}}_{QL}^T, \hat{\sigma}_{QL}\right)^T$ is obtained using Newton-Raphson iteration as $\hat{\boldsymbol{\theta}}_{QL}(t+1) = \hat{\boldsymbol{\theta}}_{QL}(t) + \left[\sum_{i=1}^n \frac{\partial \boldsymbol{M}_{\boldsymbol{i}}'(\theta)}{\partial\theta}\boldsymbol{\Omega}_{\boldsymbol{i}}^{-1}(\theta)\frac{\partial \boldsymbol{M}_{\boldsymbol{i}}(\theta)}{\partial\theta}\right]_{(t)}^{-1}\left[\sum_{i=1}^n \frac{\partial \boldsymbol{M}_{\boldsymbol{i}}'(\theta)}{\partial\theta}\boldsymbol{\Omega}_{\boldsymbol{i}}^{-1}(\theta)[\boldsymbol{S}_{\boldsymbol{i}} - \boldsymbol{M}_{\boldsymbol{i}}(\theta)]\right]$ with covariance V where $\hat{V}\left(\hat{\boldsymbol{\theta}}_{QL}\right) = \left[\sum_{i=1}^n \frac{\partial \boldsymbol{M}_{\boldsymbol{i}}'(\theta)}{\partial\theta}\boldsymbol{\Omega}_{\boldsymbol{i}}^{-1}(\theta)\frac{\partial \boldsymbol{M}_{\boldsymbol{i}}(\theta)}{\partial\theta}\right]^{-1}$. These GQL estimators are consistent and efficient.[23] A specification of the GQL model is important as consistency of the regression parameter estimates depends on correctly specifying the link function and the efficiency depends on a correctly specified variance function.

3. Modified Generalized Quasi-likelihood Models using the Canonical Mean-Variance Parameterization

3.1 Canonical Parameterization

For a random variable Y, we consider describing the variance in terms of two parameters including a dispersion parameter ψ and power parameter λ. Tsou[9] suggested the mean-variance relationship, $\psi\mu^\lambda$, which introduced additional flexibility compared to the one parameter form. However, this form is limited to distributions with a variance power relationship such as the Poisson or gamma distributions ($\lambda = 1$ or $\lambda = 2$, respectively). We generalize this power parameterization through the canonical parameter. Consider the canonical parameter θ through its derivative of the inverse link function h, where $h = g^{-1}$. Then, we propose the general mean-variance relationship as

$$var(Y) = \psi[h'(g(\mu))]^{\lambda} = \psi[h'(\theta)]^{\lambda}$$

where $h'(\theta)$ is the first derivative of the inverse of the canonical link and ψ and λ are parameters measuring the overdispersion, $\psi > 0$. This form of the mean-variance relationship is distinct as it is applicable to members of the exponential family of distributions and provides general flexibility in describing the mean-variance relationship.

For example, if Y follows a Poisson distribution with natural parameter α, then the canonical parameter is $\theta = log(\alpha)$ and the corresponding mean-variance parameter relation is

$$var(Y) = \psi[h'(\theta)]^{\lambda} = \psi\alpha^{\lambda}.$$

If Y follows a binomial distribution with natural parameters m and p and canonical parameter $\theta = logit(p)$, then the canonical mean-variance parameter relation is

$$var(Y) = \psi[h'(\theta)]^{\lambda} = \psi\left[\frac{exp(\theta)}{(1 + exp(\theta))^2}\right]^{\lambda} = \psi[p(1-p)]^{\lambda},$$

as $p = e^{\theta}\left(1 + e^{\theta}\right)^{-1}$. Thus, the dispersion parameter ψ and power parameter λ are a means of adjusting for deviation from the assumed distributional properties. Values of the overdispersion parameters ψ and λ different than 1 indicate violations in the variance assumption. While the binomial does not have a natural power relationship between the mean and variance, the relationship is tractable in its power form when based on the canonical parameter. We measure the deviations and consider distributions that are not fully identified but belong to the quasi-exponential family. Such distributions are identified based on the scale of the responses which is robust to misspecification.

3.2 Estimation of ψ and λ

Let $\hat{\boldsymbol{\gamma}}_{GMM}$ be an estimator for a vector of parameters $\boldsymbol{\gamma} = (\psi, \lambda)^T$ that minimizes the quadratic objective function $f_n(\boldsymbol{\gamma})^T W_n f_n(\boldsymbol{\gamma})$ where $f_n(\boldsymbol{\gamma})$ is a function of the vector of the sample moment conditions, and $\boldsymbol{W}_n$ is a symmetric, positive definite weight matrix of dimension n.[26, 27] Then,

$$\hat{\boldsymbol{\gamma}}_{GMM} = argmin_{\beta}\left\{f_n(\boldsymbol{\gamma})^T \boldsymbol{W}_n f_n(\boldsymbol{\gamma})\right\} \tag{3.1}$$

is a generalized method of moments estimator for $\boldsymbol{\gamma}$ which minimizes the objective function. Thus, the GMM estimators of the parameters ψ and λ, are obtained from the population moment conditions

$$E\left(h'(\theta_i)\left(var(y_i) - \psi[h'(\theta_i)]^{\lambda}\right)\right) = 0 \tag{3.2a}$$

$$E\left(h'(\theta_i)^2\left(var(y_i) - \psi[h'(\theta_i)]^\lambda\right)\right) = 0 \tag{3.2b}$$

where $h'(\theta_i)$ is the first derivative of the inverse link function and $var(y_i)$ is an empirical estimate of the variance based on the squared residual, $(y_i - \mu_i)^2$. Equating the moment conditions and an empirical estimate of $f_n(\boldsymbol{\gamma})$ results in

$$\frac{1}{n}\sum_{i=1}^{n} f(y_i, \psi, \lambda) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} h'(\theta_i)\left(var(y_i) - \psi[h'(\theta_i)]^\lambda\right) \\ \frac{1}{n}\sum_{i=1}^{n} h'(\theta_i)^2\left(var(y_i) - \psi[h'(\theta_i)]^\lambda\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

We make use of a two-step GMM approach, with an identity weight matrix in the first step. In the second step, the weights are selected as an estimate of the optimal weight matrix for GMM as

$$\widehat{\boldsymbol{W}}_n = \left[\frac{1}{n}\sum_{i=1}^{n} f\left(y_i, \hat{\psi}, \hat{\lambda}\right) f\left(y_i, \hat{\psi}, \hat{\lambda}\right)^T\right]^{-1}$$

where $\hat{\psi}$ and $\hat{\lambda}$ are mean-variance relationship parameter estimates from the first step.[28] Thus, the vector of GMM estimates for the mean-variance relation parameters $\hat{\boldsymbol{\gamma}}_{GMM}$ minimizes the quadratic objective function $f_n(\boldsymbol{\gamma})^T \boldsymbol{W}_n f_n(\boldsymbol{\gamma})$. The generalized method of moments approach is flexible and estimates both parameters (ψ, λ) simultaneously.

An alternative approach is to fix one parameter at a time and estimate the second parameter using one moment condition. Thus, an extension of GMM is to make use of additional moment conditions, such as

$$f_3(y_i, \psi, \lambda) = h'(\theta_i)^3\left(var(y_i) - \psi[h'(\theta_i)]^\lambda\right) \tag{3.3}$$

to estimate the parameters. The additional moment condition improves the asymptotic efficiency, although there is a possibility of small sample bias.[29]

To identify the mean-variance relation in clustered data, consider y_{ij}, the j^{th} observation in the i^{th} cluster, $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$, with mean μ_{ij} which is related to k covariates and random effect α_i through the link function g such that

$$E\left(Y_{ij}\right) = \mu_{ij}$$

and

$$g\left(\mu_{ij}\right) = \eta_{ij} = x_{ijk}^T\beta + \alpha_i$$

where the random effect $\boldsymbol{\alpha}_i$ represents the variation between clusters such that $\alpha_i \sim N(0, \sigma^2)$. Further, let $\xi_i = \alpha_i / \sigma$, such that $\xi_i \sim N(0, 1)$ so the linear predictor reduces to

$$g(\mu_{ij}) = x_{ijk}^T \beta + \sigma \xi_i \,.$$

The general mean-variance relationship, obtained for the data across all the clusters, is

$$var(Y) = \psi[h'(\theta)]^{\lambda},$$

where θ is the canonical parameter, $h'(\theta)$ is the first derivative of the inverse canonical link, and ψ and λ are the variance parameters estimated from the data, (3.2a) and (3.2b). The estimates of ψ and λ indicate the strength of the clustering (variance of the random effect).

The parameters ψ and λ are essential in defining the variance and identifying deviations from the theoretical values in a known distributional mean-variance relation. We obtain GMM estimators using the first and second moments based on the fact that the distribution is a member of the quasi-exponential family.[30] We do not require complete distributional assumptions, as is required with maximum likelihood estimators, and the estimates are obtainable even when likelihood methods are computationally burdensome.[26] The GMM estimators for ψ and λ are consistent and asymptotically normal.[31]

3.3 Inference for ψ and λ

Assume that the data come from a quasi-exponential family. The sample moments are asymptotically normally distributed, so we have

$$\sqrt{n}(f_n(\hat{\boldsymbol{\gamma}})) \xrightarrow{d} N(0, \boldsymbol{\Delta}),$$

with the asymptotic variance $\boldsymbol{\Delta} = E\left[f(y, \boldsymbol{\gamma}^*) f(y, \boldsymbol{\gamma}^*)^T\right]$ where $\boldsymbol{\gamma}^*$ is the estimate of the parameters of interest.[30] For the mean-variance relationship parameters ψ and λ, the GMM estimator $\hat{\boldsymbol{\gamma}}_{GMM} = \left(\hat{\psi}_{GMM}, \hat{\lambda}_{GMM}\right)$ has the asymptotic covariance

$$var(\hat{\boldsymbol{\gamma}}_{GMM}) = \boldsymbol{V}_{GMM} = \frac{1}{n}\left[\boldsymbol{\Gamma}^T \boldsymbol{W} \Gamma\right]^{-1} \boldsymbol{\Gamma}^T \boldsymbol{W} \Delta W \Gamma \left[\boldsymbol{\Gamma}^T \boldsymbol{W} \Gamma\right]^{-1}$$

where $\boldsymbol{W}$ is a specified weight matrix and $\boldsymbol{\Gamma}$ is the expected value of the Jacobian of population moment conditions found as

$$\boldsymbol{\Gamma} = E\left[\frac{\partial f(y, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right] = E\left[\frac{\partial f(y, \lambda, \psi)}{\partial \psi}, \frac{\partial f(y, \lambda, \psi)}{\partial \lambda}\right]^T .$$

In the optimal case, the weight matrix is selected as $\boldsymbol{W} = (\boldsymbol{\Delta})^{-1}$, so that

$$\boldsymbol{V}_{GMM} = \frac{1}{n}[\boldsymbol{\Gamma}'\boldsymbol{W}\boldsymbol{\Gamma}]^{-1}$$

resulting in asymptotically efficient GMM estimators, $\hat{\psi}_{GMM}$ and $\hat{\lambda}_{GMM}$.[26] In practice, the covariance matrix is evaluated using $\hat{\boldsymbol{\gamma}}$,

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial f(y, \hat{\boldsymbol{\gamma}})}{\partial \hat{\boldsymbol{\gamma}}}.$$

Significant overdispersion is identified through two hypothesis tests of the overdispersion parameters ψ and λ,

$$H_0 : \psi = 1,\ H_a : \psi > 1$$

and

$$H_0 : \lambda = 1,\ H_a : \lambda > 1.$$

Then the z-test statistics

$$Z_\psi = \frac{\hat{\psi} - 1}{\sqrt{\mathrm{var}(\hat{\psi})}}$$

and

$$Z_\lambda = \frac{\hat{\lambda} - 1}{\sqrt{\mathrm{var}(\hat{\lambda})}},$$

follow the standard normal distribution under the null hypothesis. Thus, a measure of the overdispersion is given based on the joint 100(1-α)% confidence intervals for ψ and λ,

$$\left(\hat{\psi}_{GMM} - z_{1-\frac{\alpha}{2}}\sqrt{V_{GMM,\psi}}, \hat{\psi}_{GMM} + z_{1-\frac{\alpha}{2}}\sqrt{V_{GMM,\psi}}\right)$$

$$\left(\hat{\lambda}_{GMM} - z_{1-\frac{\alpha}{2}}\sqrt{V_{GMM,\lambda}}, \hat{\lambda}_{GMM} + z_{1-\frac{\alpha}{2}}\sqrt{V_{GMM,\lambda}}\right)$$

where z_α is the α^{th} quantile from the standard normal distribution.[28]

3.4 Modified Generalized Quasi-likelihood Models

In this section, we propose a modified generalized quasi-likelihood model for correlated data based on the canonical parameterization. As correlated data necessitate dealing with extravariation, we rely on our two-parameter mean-variance relation. The GQL approach relies on the specification of the mean-variance relationship rather than a distributional assumption. We address the correlation through the empirical mean-variance estimates of ψ and λ.

The generalized quasi-likelihood estimating equation (2.1), with $\boldsymbol{M_i}$, $\frac{\partial \boldsymbol{M_i'}(\theta)}{\partial \theta}$, and the covariance matrix $\boldsymbol{\Omega_i}(\theta)$, is used to estimate the regression parameters $\boldsymbol{\beta}$ and the variance of the random effect $\boldsymbol{\sigma}$. Though GQL performs well and produces consistent and efficient estimators,[23] it relies on the estimate of the covariance $\boldsymbol{\Omega_i}(\theta)$. We update this estimate to incorporate extravariation based on the canonical parameterization $\psi\boldsymbol{\Omega_i}(\theta)^{\lambda}$ such that the modified quasi-likelihood estimate $\hat{\theta}_{MGQL}$ is obtained iteratively as

$$\hat{\theta}_{MGQL}(t+1) = \hat{\theta}_{MGQL}(t) + \left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{M_i'}(\theta)}{\partial \theta}\left(\psi \boldsymbol{\Omega}_i^{\lambda}(\theta)\right)^{-1}\frac{\partial \boldsymbol{M_i}(\theta)}{\partial \theta}\right]_{(t)}^{-1}$$

$$\left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{M_i}^{'}(\theta)}{\partial \theta}\left(\psi \boldsymbol{\Omega}_i^{\lambda}(\theta)\right)^{-1}[\boldsymbol{S_i} - \boldsymbol{M_i}(\theta)]\right].$$

with covariance

$$\hat{V}\left(\hat{\theta}_{MGQL}\right) = \left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{M_i'}(\theta)}{\partial \theta}\left(\psi \boldsymbol{\Omega}_i^{\lambda}(\theta)\right)^{-1}\frac{\partial \boldsymbol{M_i}(\theta)}{\partial \theta}\right]^{-1}.$$

This modification makes use of the GMM estimates of ψ and λ. For given ψ and λ, the MGQL estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ in (2.1) are unbiased. The MGQL estimators are consistent and efficient as μ_{ij} is the mean of y_{ij}.

4. Simulation Study

We simulate hierarchical binary data and evaluate the estimation of the regression parameters using the MGQL model which incorporates GMM estimates of the mean-variance parameters into the quasi-likelihood model framework, a GQL model, and a generalized linear mixed model (GLMM) over 1000 iterations. The two-level binary data contain 50 clusters with 10 observations in each cluster, with the linear predictor $\mathrm{log}it(\mu_i) = \eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \alpha_i$ where $\beta_1 = \beta_2 = 1$ and X_1 and X_2 are generated from standard normal distributions. The canonical mean-variance parameter relation under the Bernoulli distribution is $\mathrm{var}(Y) = \psi[p(1-p)]^{\lambda}$. To evaluate the performance of these regression methods under the true mean-variance relation and an overdispersed form, we consider cases where the random effects are generated under the normal distribution and the t-distribution with 4 degrees of freedom. The GLMM is fit using the default optimization techniques in the R statistical software, which utilizes Nelder-Mead for the preliminary optimization of the random effects parameters and the bobyqa optimizer from the minqa package for the final estimation of the random and fixed effects parameters.

4.1 Normally Distributed Random Effects

We evaluate hierarchical binary data with normally distributed random effects. The random intercept α_i associated with each cluster is generated from $N\left(0, \sigma^2\right)$ with σ = 0.6, 0.8, 1, 1.2, 1.4. The results are reported in Table 1. The generalized linear mixed model estimates for the standard error of $\hat{\sigma}$ were obtained using a profile likelihood approach.

The simulation results demonstrate that the MGQL approach performs well and suggests that the MGQL model recovers the true values when relying on the estimated mean-variance relationship in the covariance matrix. While the parameter estimates are similar across the three methods, the standard errors for the MGQL estimates of β_1 and β_2 are lower than the standard errors of the GQL approach across all values of σ. The MGQL approach requires slightly more iterations, on average, than the GQL approach to achieve convergence, although the two approaches require similar computation time. As expected, the GLMM performs the best among the three methods as the data are generated under this model.

4.2 t-Distributed Random Effects

We evaluate the performance of the MGQL, GQL, and GLMM for non-normally distributed random effects. The random effects α_i are generated under the t-distribution with 4 degrees of freedom, which has heavier tails than the normal distribution. The model parameter estimates and standard errors are reported in Table 2.

As seen in the previous simulation, MGQL tends to be more efficient than the GQL approach for estimates of β. The simulation results also highlight the advantage of implementing the MGQL model for overdispersed data. For small values of σ, the simulated data reflect the true mean-variance parameterization for the Bernoulli distribution and thus we see similar performance across the three methods. However, for larger values of σ where overdispersion is present (for σ = 1.2 and σ = 1.4, there was significant overdispersion in 60.0% and 62.3% of simulations, respectively), we find that the MGQL model produces improved estimates of β_1 and β_2 compared to the GQL model and GLMM. In addition, for values of $\sigma > 1$, we see that MGQL produces more efficient variance estimates. Thus, when the random effects are not normally distributed, the MGQL approach has many advantages for modeling overdispersed data compared to the GQL model and GLMM.

5. Numerical Example

The Add Health Study is a longitudinal study in the United States of adolescents in 7th through 12th grade, with information collected over four waves of interviews between 1994 and 2008.[24] The data are available on the Add Health website (http://www.cpc.unc.edu/addhealth). We fit a modified generalized quasi-likelihood model to evaluate the binary variable adolescent obesity for 2,712 adolescents in the United States. The factors associated with obesity include activity scale and feeling scale, ratings of physical activity and emotional health. The mean-variance parameter estimates are $\hat{\psi} = 3.16$ and $\hat{\lambda} = 1.80$ with standard errors 0.27 and 0.06, respectively. The estimates indicate a significant deviation from the distributional form of the mean-variance relationship (test statistics $Z_\psi = 7.94$ and $Z_\lambda = 13.57$). Thus, making use of the true mean-variance form accounts for the clustering.

The model parameter estimates and standard errors of the MGQL, GQL, and GLMM are provided (Table 3).

The covariates activity scale and feeling scale as well as the random effect are found to be significant across all three models. The regression parameter estimates for activity scale are positive, indicating that increased physical activity is associated with a lower probability of obesity. Similar estimates are produced for the GQL and GLMM approaches, while the MGQL estimate is slightly smaller $\left(\beta_{Activity\ Scale} = -1.0921\right)$. Similarly, the parameter estimates for feeling scale vary slightly although all three estimates are negative, indicating that larger values of the computed emotional health measure is associated with a lower probability of obesity. The estimates of the standard error of the random effect σ vary slightly among the three models, with $\hat{\sigma}_{MGQL} = 2.6078$, $\hat{\sigma}_{GQL} = 2.7022$, and $\hat{\sigma}_{GLMM} = 2.6900$. The random effect variance for the MGQL model is found to be the lowest of the three estimates.

6. Conclusions

It is common to assume that the variance of a random variable is a function of the mean, although it is often the case that the true variance in the data may be inflated due to underlying correlation or the hierarchical data structure. While the presence of overdispersion impacts the accuracy of statistical evaluations, the MGQL is a modeling approach that appropriately fits correlated data. The MGQL approach is flexible as it accounts for correlation through an extended representation of the covariance. The canonical parameterization is tractable in the power form for any distribution in the exponential family. Moreover, deviations in the variance can be readily identified using the proposed GMM estimators of the mean-variance parameters ψ and λ which are consistent and asymptotically normal. A simulation study demonstrated that the MGQL addresses correlation through the use of the mean-variance relationship and performs as well or better than existing methods including GQL models and GLMM, particularly for non-normally distributed random effects. The study confirmed that the MGQL retains good properties of quasi-likelihood approaches including unbiased estimates and small standard errors. In addition, we consider a numerical example to evaluate obesity data from the Add Health study. We verified that the MGQL model produced comparable results to existing models. Factors including activity scale and feeling scale were found to be negatively associated with obesity, and the MGQL model produced a lower variance estimate of the random effect.

Acknowledgements

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation [grant number DGE-1311230] and the National Institutes of Health Alzheimer's Consortium Fellowship Grant [grant number NHS0007].

References

1. Morel JG and Neerchal NK. Overdispersion Models in SAS Cary: SAS Institute Inc, 2012.
2. Cox DR. Some remarks on overdispersion. Biometrika 1983; 70: 269–274.
3. Wilson JR and Koehler KJ. Hierarchical Models for Cross-classified Overdispersed Multinomial Data. Journal of Business and Economic Statistics 1991; 9: 103–110.
4. Milanzi E, Alonso A and Molenberghs G. Ignoring overdispersion in hierarchical loglinear models: Possible problems and solutions. Statistics in Medicine 2012; 31: 1475–1482. DOI: 10.1002/sim.4482. [PubMed: 22362329]
5. McCullagh P and Nelder JA. Generalized Linear Models, 2nd *ed* London: Chapman and Hall, 1989.
6. Cox DR and Reid N. Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society Series B 1987; 49: 1–39.
7. McCullagh P and Tibshirani R. A Simple Method for the Adjustment of Profile Likelihoods. Journal of the Royal Statistical Society Series B 1990; 52: 325–344.
8. Kukush A, Malenko A and Schneeweiss H. Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models. Journal of Statistical Planning and Inference 2009; 139: 3461–3472.
9. Tsou T-S. Determining the mean-variance relationship in generalized linear models—A parametric robust way. Journal of Statistical Planning and Inference 2011; 141: 197–203.
10. Pack SE. Hypothesis Testing for Proportions with Overdispersion. Biometrics 1986; 42: 967–972. [PubMed: 3814737]
11. Dean C Testing for overdispersion in Poisson and Binomial regression models. Journal of the American Statistical Association 1992; 87: 451–457.
12. Xiang L, Lee AH, Yau KKW, et al. A score test for overdispersion in zero-inflated Poisson mixed regression model. Statistics in Medicine 2007; 26: 1608–1622. [PubMed: 16794991]
13. Yang Z, Hardin JW and Addy CL. A note on Dean's overdispersion test. Journal of Statistical Planning and Inference 2009; 139: 3675–3678.
14. Wilson JR and Lorenz KA. Modeling Binary Correlated Responses using SAS, SPSS and R New York: Springer, 2015.
15. Liang K-Y and Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73: 13–22.
16. Lee Y and Nelder JA. Two ways of modelling overdispersion in non-normal data. *Journal of the Royal* Statistical Society Series C 2000; 49: 591–598.
17. Breslow NE and Clayton DG. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 1993; 88: 9–25.
18. Smyth GK. Generalized linear models with varying dispersion. Journal of the Royal Statistical Society Series B 1989; 51: 47–60.
19. Smyth GK and Verbyla AP. Adjusted likelihood methods for modelling dispersion in generalized linear models. Environmetrics 1999; 10: 695–709.
20. Lee Y and Nelder JA. Double hierarchical generalized linear models. *Journal of the Royal Statistical* Society Series C 2006; 55: 139–185.
21. Wedderburn R Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 1974; 61: 439–447.
22. Wang B and Wilson JR. Comparative GMM and GQL logistic regression models on hierarchical data. Journal of Applied Statistics 2017; 45: 409–425.
23. Sutradhar BC. On Exact Quasilikelihood Inference in Generalized Linear Mixed Models. Sankhy-a: The Indian Journal of Statistics 2004; 66: 263–291.

24. Harris KM, Halpern C, Whitsel E, et al. The National Longitudinal Study of Adolescent to Adult Health: Research Design http://www.cpc.unc.edu/projects/addhealth/design 2009.
25. Dobson AJ and Barnett AG. An Introduction to Generalized Linear Models, Third Edition New York: CRC Press, 2008.
26. Zsohar P Short Introduction to the Generalized Method of Moments. Hungarian Statistical Review Special Number 16 2012; 90: 150–170.
27. Lalonde TL, Wilson JR and Yin J. GMM logistic regression models for longitudinal data with time-dependent covariates and extended classifications. Statistics in Medicine 2014; 33: 4756–4769. [PubMed: 25130989]
28. Imbens GW and Spady R. Confidence intervals in generalized method of moments models. Journal of Econometrics 2002; 107: 87–98.
29. Donald SG, Imbens GW and Newey WK. Choosing instrumental variables in conditional moment restriction models. Journal of Econometrics 2009; 152: 28–36.
30. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982; 50: 1029–1054.
31. Jiang J Empirical method of moments and its applications. Journal of Statistical Planning and Inference 2003; 115: 69–84.

Table 1:

Model Fit Simulation Results for Normally Distributed Random Effects

		β_1		β_2		σ		Iterations
		Est	**SE**	**Est**	**SE**	**Est**	**SE**	
$\sigma = 0.6$	MGQL	1.0083	0.1228	1.0181	0.1231	0.5687	0.2177	5.5
	GQL	1.0053	0.1319	1.0120	0.1320	0.5837	0.2019	5.2
	GLMM	1.0040	0.1316	1.0110	0.1318	0.5706	0.1872	-
$\sigma = 0.8$	MGQL	1.0093	0.1243	1.0241	0.1247	0.7822	0.1880	4.5
	GQL	1.0045	0.1344	1.0159	0.1348	0.7948	0.1814	4.0
	GLMM	1.0035	0.1327	1.0149	0.1331	0.7790	0.1869	-
$\sigma = 1$	MGQL	1.0115	0.1281	1.0237	0.1284	0.9841	0.1924	4.5
	GQL	1.0047	0.1372	1.0137	0.1375	1.0014	0.1916	4.0
	GLMM	1.0041	0.1286	1.0131	0.1288	0.9832	0.1943	-
$\sigma = 1.2$	MGQL	1.0103	0.1333	1.0225	0.1336	1.1837	0.2067	4.6
	GQL	1.0025	0.1401	1.0126	0.1404	1.2044	0.2096	4.1
	GLMM	1.0023	0.1220	1.0125	0.1222	1.1835	0.2108	-
$\sigma = 1.4$	MGQL	1.0148	0.1402	1.0253	0.1405	1.3951	0.2280	4.7
	GQL	1.0078	0.1437	1.0146	0.1439	1.4120	0.2326	4.2
	GLMM	1.0080	0.1137	1.0147	0.1139	1.3874	0.2335	-

Table 2:

Model Fit Simulation Results for t-Distributed Random Effects

		β_1		β_2		σ		Iterations
		Est	**SE**	**Est**	**SE**	**Est**	**SE**	
$\sigma = 0.6$	MGQL	1.0185	0.1245	1.0141	0.1243	0.7653	0.1922	5.0
	GQL	1.0148	0.1347	1.0099	0.1345	0.7831	0.1818	4.0
	GLMM	1.0136	0.1324	1.0089	0.1320	0.7671	0.1883	-
$\sigma = 0.8$	MGQL	1.0150	0.1287	1.0071	0.1283	1.0047	0.1953	4.7
	GQL	1.0143	0.1379	1.0041	0.1374	1.0214	0.1935	4.1
	GLMM	1.0143	0.1281	1.0041	0.1275	1.0075	0.1977	-
$\sigma = 1$	MGQL	1.0134	0.1354	1.0119	0.1352	1.2435	0.2136	4.9
	GQL	1.0118	0.1414	1.0111	0.1413	1.2542	0.2151	4.2
	GLMM	1.0127	0.1190	1.0120	0.1191	1.2417	0.2194	-
$\sigma = 1.2$	MGQL	1.0104	0.1427	1.0090	0.1425	1.4770	0.2385	4.7
	GQL	1.0102	0.1450	1.0093	0.1449	1.4821	0.2414	4.3
	GLMM	1.0119	0.1207	1.0110	0.1205	1.4698	0.2466	-
$\sigma = 1.4$	MGQL	1.0128	0.1507	1.0050	0.1503	1.7140	0.2705	4.8
	GQL	1.0135	0.1490	1.0076	0.1487	1.7109	0.2716	4.4
	GLMM	1.0155	0.1328	1.0096	0.1327	1.6969	0.2770	-

Table 3:

Parameter estimates and standard errors for adolescent obesity data

		$\beta_{Activity\ Scale}$	$\beta_{Feeling\ Scale}$	σ
MGQL	Estimate	−1.0921	−0.6758	2.6078
	Std. Error	0.0326	0.0650	0.0794
GQL	Estimate	−1.3458	−0.5041	2.7022
	Std. Error	0.0449	0.0718	0.0974
GLMM	Estimate	−1.3438	−0.6527	2.6900
	Std. Error	0.0471	0.0726	0.0959