Generalized Estimating Equations using the new R package glmtoolbox

by L.H. Vanegas, L.M. Rondón, and G.A. Paula

Abstract This paper introduces a very comprehensive implementation, available in the new `R` package `glmtoolbox`, of a very flexible statistical tool known as Generalized Estimating Equations (GEE), which analyzes cluster correlated data utilizing marginal models. As well as providing more built-in structures for the working correlation matrix than other GEE implementations in `R`, this GEE implementation also allows the user to: (1) compute several estimates of the variance-covariance matrix of the estimators of the parameters of interest; (2) compute several criteria to assist the selection of the structure for the working-correlation matrix; (3) compare nested models using the Wald test as well as the generalized score test; (4) assess the goodness-of-fit of the model using Pearson-, deviance- and Mahalanobis-type residuals; (5) perform sensibility analysis using the global influence approach (that is, dfbeta statistic and Cook's distance) as well as the local influence approach; (6) use several criteria to perform variable selection using a hybrid stepwise procedure; (7) fit models with nonlinear predictors; (8) handle dropout-type missing data under MAR rather than MCAR assumption by using observation-specific or cluster-specific weighted methods. The capabilities of this GEE implementation are illustrated by analyzing four real datasets obtained from longitudinal studies.

1 Introduction

The Generalized Estimating Equations (GEE), proposed by Liang and Zeger (1986), extend the theoretical framework of the Generalized Least Squares (GLS) by allowing the variance of the response variable distribution to be proportional to a known function of its mean, resulting thus in a very flexible statistical tool for the analysis of heteroskedastic discrete and continuous cluster correlated data. Unlike conditional models such as random-effect models, the GEE approach is based on marginal models. In addition, and according to Lipsitz and Fitzmaurice (2008), GEE can also be regarded as a multivariate generalization of the quasi-likelihood approach to Generalized Linear Models (GLMs) introduced by Wedderburn (1974). The main advantage of GEE over other approaches to analyzing cluster correlated data lies in that this methodology does not require the full specification of the multivariate distribution of the (discrete or continuous) response vector measured on each subject or cluster, reducing the possibility of model misspecification. Indeed, GEE just requires the following:

- Specification of a variance function, which describes the mechanism of heteroscedasticity (if there is any), that is, it describes the way in which the variance of the response variable distribution is assumed to be dependent on its mean.
- Specification of a regression structure, very similar to that described in the theoretical framework of the GLMs (see, for instance, McCullagh and Nelder (1989)), that includes a link function and a linear predictor, which describe the way the mean of the response variable distribution is assumed to be dependent on some continuous and/or discrete regressors.
- Specification of a correlation matrix structure. This matrix describes the dynamic of the linear association between the different measurements of the response variable performed on the same subject or cluster.

This paper introduces the package `glmtoolbox`, which, besides providing more built-in structures for the working correlation matrix than other GEE implementations available in `R`, has several features, including: (1) compute several estimates of the variance-covariance matrix of the estimators of the parameters of interest; (2) compute several criteria to assist the selection of the structure for the working correlation matrix; (3) compare nested models using the Wald test as well as the generalized score test; (4) assess the goodness-of-fit of the model using Pearson-, deviance- and Mahalanobis-type residuals; (5) perform sensibility analysis using the global influence approach (that is, dfbeta statistic and Cook's distance) as well as the local influence approach; (6) use several criteria to perform variable selection using a hybrid stepwise procedure; (7) fit models with nonlinear predictors; (8) handle dropout-type missing data under MAR rather than MCAR assumption by using observation-specific or cluster-specific weighted methods. The rest of this paper is organized as follows: in Section 2 the main features of the GEE model setup are described; in Section 3 the main features of the implementation of GEE in the package `glmtoolbox` are described and compared with those available in the packages `gee` (Carey, 2022), `geepack` (Yan, 2002; Højsgaard et al., 2005) and `geeM` (McDaniel et al., 2013), which are the most widely used packages in `R` to analyze cluster correlated data using GEE; in Section 4 the capabilities of this implementation are illustrated by analyzing four real datasets obtained from longitudinal studies.

2 Generalized Estimating Equations

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{i\,n_i})^\top$ for $i = 1, \ldots, n$ be the multivariate response of interest measured on n subjects or clusters, which are assumed to be realizations of independent random vectors denoted here by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ij}, \ldots, Y_{i\,n_i})^\top$ for $i = 1, \ldots, n$, where n_i represents the size of the ith cluster or the number of measurements performed on the ith subject. So, the total number of observations is $N = n_1 + \ldots + n_n$. The random variables associated with the ith subject or cluster, given by Y_{ij} for $j = 1, \ldots, n_i$, are assumed to satisfy the following:

$$\mathrm{Var}(Y_{ij}) = \frac{\phi}{\omega_{ij}} \mathrm{V}(\mu_{ij}) \qquad \text{and} \qquad \mathrm{Corr}(Y_{ij}, Y_{ik}) = r_{jk}(\boldsymbol{\rho}),$$

where $\mu_{ij} = \mathrm{E}(Y_{ij})$, $\phi > 0$ is the dispersion parameter, $\omega_{ij} > 0$ are known weights typically specified to be 1, $\mathrm{V}(\mu) > 0$ is the variance function, and $r_{jk}(\boldsymbol{\rho})$ is the Pearson's linear correlation coefficient, which is assumed to be dependent just on j, k and the unknown nuisance parameter vector denoted here by $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)^\top$. In addition, μ_{ij} is assumed to be dependent on a vector of p continuous and/or discrete regressors, denoted here by $(x_{1ij}, \ldots, x_{pij})$, in the following way:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^\top \boldsymbol{\beta}, \tag{1}$$

where $g(\mu)$ is a strictly monotone and twice-differentiable known function better known as link function, $\mathbf{x}_{ij} = (1, x_{1ij}, \ldots, x_{pij})^\top$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is the interest parameter vector.

According to Liang and Zeger (1986), the estimate of $\boldsymbol{\beta}$, denoted here by $\hat{\boldsymbol{\beta}}$, reduces to the solution to the $(p+1)$ equations given by $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, where

$$\mathbf{U}(\boldsymbol{\beta}) = \phi^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{K}_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \phi^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{W}_i \mathbf{K}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \phi^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{K}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \tag{2}$$

in which $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^\top$, $\mathbf{W}_i = \mathbf{K}_i \mathbf{V}_i^{-1} \mathbf{K}_i$, $\mathbf{K}_i = \mathrm{diag}\{1/g'(\mu_{i1}), \ldots, 1/g'(\mu_{in_i})\}$, $\mathbf{V}_i = \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i \mathbf{A}_i^{\frac{1}{2}}$, $\mathbf{A}_i = \mathrm{diag}\{\mathrm{V}(\mu_{i1})/\omega_{i1}, \ldots, \mathrm{V}(\mu_{in_i})/\omega_{in_i}\}$, $\mathbf{R}_i$ is a square matrix whose (j,k)th entry is $r_{jk}(\boldsymbol{\rho})$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)^\top$, $\mathbf{W} = \mathrm{blockdiag}\{\mathbf{W}_1, \ldots, \mathbf{W}_n\}$, $\mathbf{K} = \mathrm{blockdiag}\{\mathbf{K}_1, \ldots, \mathbf{K}_n\}$, $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_n^\top)^\top$. Moreover, the estimate of ϕ may be written as follows:

$$\hat{\phi} = \frac{1}{N - p - 1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\mathrm{V}(\hat{\mu}_{ij})/\omega_{ij}},$$

where $\hat{\mu}_{ij} = g^{-1}(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}})$. If the model for the mean (μ) is correctly specified, then, under certain regularity conditions, $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ and its distribution is such that (Liang and Zeger, 1986):

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathrm{Var}(\hat{\boldsymbol{\beta}})),$$

where

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \lim_{n \to \infty} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{W} \mathbf{X} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{W}_i \mathbf{K}_i^{-1} \mathrm{Var}(\mathbf{Y}_i) \mathbf{K}_i^{-1} \mathbf{W}_i \mathbf{X}_i \right) \left(\frac{1}{n} \mathbf{X}^\top \mathbf{W} \mathbf{X} \right)^{-1}.$$

Therefore, if the mean model is correctly specified, then $\hat{\boldsymbol{\beta}}$ remain consistent and asymptotically normal distributed regardless of whether or not the correlation matrix structure is correctly specified. Indeed, if the structure of the correlation matrix is also correctly specified, that is, if $\mathrm{Var}(\mathbf{Y}_i) = \phi \mathbf{V}_i$ for $i = 1, \ldots, n$, then $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ reduces to

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \lim_{n \to \infty} \phi \left(\frac{1}{n} \mathbf{X}^\top \mathbf{W} \mathbf{X} \right)^{-1}.$$

3 R package `glmtoolbox`

The function `glmgee()` is the GEE solver available in the package `glmtoolbox`. That function includes the typical arguments present in regression routines such as `lm()` and `glm()`, that is, it includes arguments such as `formula`, `weights`, `start`, `data` and `subset`. In addition, the objects generated by the function `glmgee()` are associated with the typical extraction methods such as `summary()`, `print()`, `coef()`, `vcov()`, `fitted()`, `confint()`, `anova()`, `residuals()`, `predict()`, `leverage()`, `dfbeta()` and `cooks.distance()`. Next, the main features of the implementation of GEE in `glmtoolbox` are described and compared with those available in the packages gee, geepack and `geeM`.

3.1 Link and variance functions

The available options for the link ($g(\mu)$) and variance function ($\mathrm{V}(\mu)$) in the routine `glmgee()` are the following:

`family`	$\mathrm{V}(\mu)$	$g(\mu)$
`gaussian`	1	`inverse` $\left(\mu^{-1}\right)$, `identity` (μ), `log` $(\log(\mu))$
`binomial`	$\mu(1-\mu)$	`logit` $\left(\log\left(\frac{\mu}{1-\mu}\right)\right)$, `cloglog` $(\log(-\log(1-\mu)))$, `probit` $\left(\Phi^{-1}(\mu)\right)$, `cauchit` $\left(\tan\left(\frac{\pi}{2}(2\mu-1)\right)\right)$
`poisson`	μ	`sqrt` $\left(\mu^{\frac{1}{2}}\right)$, `identity` (μ), `log` $(\log(\mu))$
`Gamma`	μ^2	`inverse` $\left(\mu^{-1}\right)$, `identity` (μ), `log` $(\log(\mu))$
`inverse.gaussian`	μ^3	`1/mu^2` $\left(\mu^{-2}\right)$, `inverse` $\left(\mu^{-1}\right)$, `identity` (μ), `log` $(\log(\mu))$
`negative.binomial(`θ`)`[1]	$\mu(1+\mu/\theta)$	`log` $(\log(\mu))$, `identity` (μ), `sqrt` $\left(\mu^{\frac{1}{2}}\right)$
`tweedie(`θ,γ`)`[2]	μ^θ	$\log(\mu)$ if $\gamma=0$ and μ^γ if $\gamma\neq 0$

1 function available in package `MASS`
2 function available in package `statmod`

Moreover, new families and new link functions may be defined by the user as described on the help page of the routine `glm()`. The variance functions $\mathrm{V}(\mu)=\mu^3$ and $\mathrm{V}(\mu)=\mu(1+\mu/\theta)$ are not available in gee.

3.2 Estimating algorithm

The $(p+1)$ equations given by $\mathbf{U}(\hat{\boldsymbol{\beta}})=\mathbf{0}$ may be solved using the following algorithm:

Step 0: Start the counter at $t=0$; set the tolerance limit, $\epsilon>0$; set the maximum number of iterations, $n_{\max}$; and set the initial value for $\boldsymbol{\beta}$, say $\boldsymbol{\beta}^{[0]}$.

Step 1: Compute $\boldsymbol{\rho}^{[t]}$ from the Pearson's residuals evaluated at $\boldsymbol{\beta}^{[t]}$, denoted here by $r_{ij}^{[t]}$.

Step 2: Compute $\boldsymbol{\beta}^{[t+1]}=\boldsymbol{\beta}^{[t]}+[\mathbf{K}(\boldsymbol{\beta}^{[t]})]^{-1}\mathbf{U}(\boldsymbol{\beta}^{[t]})=(\mathbf{X}^\top\mathbf{W}^{[t]}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{[t]}\tilde{\mathbf{y}}^{[t]}$.

Step 3: Compute $\delta^{(t+1)}=\delta(\boldsymbol{\beta}^{[t]},\boldsymbol{\beta}^{[t+1]})$.

Step 4: Update the counter by $t=t+1$.

Step 5: Repeat Steps 1,2,3 and 4 until $\delta^{(t)}<\epsilon$ or $t>n_{\max}$.

Step 6: If $\delta^{(t)}<\epsilon$, then $\hat{\boldsymbol{\beta}}$ is defined to be $\boldsymbol{\beta}^{[t]}$. Otherwise, convergence was not achieved.

Note that,

- $\boldsymbol{\beta}^{[0]}$ is specified to be the estimate of $\boldsymbol{\beta}$ in the GLM under which the random variables Y_{ij} for $i=1,\ldots,n$ and $j=1,\ldots,n_i$ are assumed to be independent. This may be easily obtained by using the function `glm()`. However, the starting value, $\boldsymbol{\beta}^{[0]}$, also may be supplied by the user with the argument `start` of the function `glmgee()`.
- $r_{ij}^{[t]}=\dfrac{y_{ij}-\mu_{ij}^{[t]}}{\sqrt{\phi^{[t]}\,\mathrm{V}(\mu_{ij}^{[t]})/\omega_{ij}}}$ for $i=1,\ldots,n$ and $j=1,\ldots,n_i$, with $\phi^{[t]}=\dfrac{1}{N-p-1}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n_i}\dfrac{(y_{ij}-\mu_{ij}^{[t]})^2}{\mathrm{V}(\mu_{ij}^{[t]})/\omega_{ij}}$.
- $\delta(\mathbf{a},\mathbf{b})$ is a non-negative and strictly increasing function of the "difference" between the vectors $\mathbf{a}=(a_1,\ldots,a_{p+1})^\top$ and $\mathbf{b}=(b_1,\ldots,b_{p+1})^\top$. For instance, $\delta(\mathbf{a},\mathbf{b})=||\mathbf{b}-\mathbf{a}||_r$ or $\delta(\mathbf{a},\mathbf{b})=||(\mathbf{b}-\mathbf{a})^*||_r$, where $||\mathbf{a}||_r=\left(|a_1|^r+\ldots+|a_{p+1}|^r\right)^{1/r}$ for any $r\geq 1$, $||\mathbf{a}||_\infty=\max\{|a_1|,\ldots,|a_{p+1}|\}$ and $(\mathbf{b}-\mathbf{a})^*:=\left((b_1-a_1)/|a_1|,\ldots,(b_{p+1}-a_{p+1})/|a_{p+1}|\right)$. The comparison criterion in the function `glmgee()` is $\delta(\mathbf{a},\mathbf{b})=||(\mathbf{b}-\mathbf{a})^*||_\infty=\max\{|b_1-a_1|/|a_1|,\ldots,|b_{p+1}-a_{p+1}|/|a_{p+1}|\}$. In addition, the values of the tolerance limit, ϵ, and the maximum number of iterations, $n_{\max}$, may be specified in the function `glmgee()` via its arguments `toler` and `maxit`, respectively. By default, `toler` $=0.00001$ and `maxit` $=50$. If `trace=TRUE` is specified in the function `glmgee()`, then the values of $\delta(\boldsymbol{\beta}^{[t]},\boldsymbol{\beta}^{[t+1]})$ are printed until convergence is reached or the maximum number of iterations is exceeded.

- $\mathbf{K}(\boldsymbol{\beta}) = \mathrm{E}\left(\dfrac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}\right) = \phi^{-1}\sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{W}_i \mathbf{X}_i = \phi^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu})$. Therefore, $\hat{\boldsymbol{\beta}}$ may be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top \hat{\mathbf{W}}\,\hat{\tilde{\mathbf{y}}}$, in which $\hat{\mathbf{W}}$ and $\hat{\tilde{\mathbf{y}}}$ represent to $\mathbf{W}$ and $\tilde{\mathbf{y}}$ evaluated at $\hat{\boldsymbol{\beta}}$, respectively. Thus, $\hat{\boldsymbol{\beta}}$ can be regarded as the GLS estimate of $\boldsymbol{\beta}$ in a linear model such that $\mathrm{E}(\tilde{\mathbf{Y}}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{Var}(\tilde{\mathbf{Y}}) = \sigma^2\hat{\mathbf{W}}^{-1}$, with $\hat{\tilde{\mathbf{y}}}$ being the observed value of the random vector $\tilde{\mathbf{Y}}$.

The package `glmtoolbox` also includes an extracting method, named `estequa()`, associated with the objects generated by the function `glmgee()`, which allows the user to verify if, actually, the parameter estimates satisfy the generalized estimating equations, that is, the extracting method `estequa()` provides the value of the vector $\mathbf{U}(\boldsymbol{\beta})$ evaluated at the parameter estimates and the observed data.

3.3 Structures for the working-correlation matrix

The available options for the structure of the working correlation matrix in the function `glmgee()` are the following:

- `corstr=“Independence”`:

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{if } j \neq k \end{cases}$$

- `corstr=“Exchangeable”`:

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1, & \text{if } j = k, \\ \rho, & \text{if } j \neq k, \end{cases} \qquad \text{and} \qquad \rho^{[t]} = \frac{1}{M - p - 1}\sum_{i=1}^{n}\sum_{j<k} r_{ij}^{[t]}\, r_{ik}^{[t]},$$

  where $M = \frac{1}{2}\sum_{i=1}^{n} n_i(n_i - 1)$.

- `corstr=“AR-M-dependent(m)”`:
  If $m = 1$, then the values of $\mathrm{Corr}(Y_{ij}, Y_{ik})$ become

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1, & \text{if } j = k, \\ \rho^{|j-k|}, & \text{if } j \neq k, \end{cases} \qquad \text{and} \qquad \rho^{[t]} = \frac{1}{M - p - 1}\sum_{i=1}^{n}\sum_{j=1}^{n_i - 1} r_{ij}^{[t]}\, r_{i,j+1}^{[t]},$$

  where $M = \sum_{i=1}^{n} (n_i - 1)$.

- `corstr=“Stationary-M-dependent(m)”`:

$$\mathrm{Corr}(Y_{ij}, Y_{i,j+l}) = \begin{cases} 1, & \text{if } t = 0, \\ \rho_l, & \text{if } l = 1, \ldots, m, \\ 0, & \text{if } l > m, \end{cases} \qquad \text{and} \qquad \rho_l^{[t]} = \frac{1}{M_l - p - 1}\sum_{i=1}^{n}\sum_{j=1}^{n_i - l} r_{ij}^{[t]}\, r_{i,j+l}^{[t]},$$

  where $M_l = \sum_{i=1}^{n} (n_i - l)$.

- `corstr=“Non-Stationary-M-dependent(m)”`:

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1, & \text{if } j = k, \\ \rho_{jk}, & \text{if } 0 < |j - k| \leq m, \\ 0, & \text{if } |j - k| > m, \end{cases} \qquad \text{and} \qquad \rho_{jk}^{[t]} = \frac{1}{n - p - 1}\sum_{i=1}^{n} r_{ij}^{[t]}\, r_{ik}^{[t]},$$

- `corstr=“Unstructured”`:

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1, & \text{if } j = k, \\ \rho_{jk}, & \text{if } j \neq k, \end{cases} \qquad \text{and} \qquad \rho_{jk}^{[t]} = \frac{1}{n - p - 1}\sum_{i=1}^{n} r_{ij}^{[t]}\, r_{ik}^{[t]},$$

- `corstr=“User-defined”`:
  Supplied by the user at the argument `corr`.

In geepack the structure `Stationary-M-dependent` is not available. Furthermore, `Non-Stationary-M-dependent` and `AR-M-dependent` (for $m > 1$) structures are not available in geepack nor in geeM.

3.4 Missing values

The rows of the data set with the same value in the variable specified by the argument `id` of the function `glmgee()` are assumed to belong to the same cluster regardless of whether they are located in consecutive rows. If the data are longitudinal, that is, if the data consist of measurements performed on the same subject/cluster but at different time points, then, by default, the function `glmgee()` assumes that the rows belonging to the same subject/cluster are ordered in time. However, the function `glmgee()` allows specifying, via its argument `waves`, an integer-valued vector, which by default is set to be $1, \ldots, n_i$, with the time points of the rows corresponding to each subject/cluster.

In longitudinal data, in which the structures for the working correlation matrix such as `AR-M-dependent`, `Stationary-M-dependent`, `Non-Stationary-M-dependent` and `Unstructured` become meaningful, the missing values may be present in one of the following ways:

- Missing values are located at the first time points.
- Intermittent missing values, that is, missing values intermixed with non-missing values in time.
- Missing values located at the last time points, also known as dropouts.

Similar to the packages `geepack` and `geeM`, the GEE solver in the package `glmtoolbox` allows the user to specify, via its argument `waves`, the way in which the missing values occurred, that is, it allows the user to specify an integer-valued vector with the time points of the non-missing values. By default, `waves` is set to be $1, \ldots, n_i$, meaning the missing values, if any, occurred at the last time points. The missing-data pattern is assumed to be *Missing Completely At Random* (MCAR) (see, for instance, Laird (1988)). Statistical inferences based on the GEE approach under the presence of missing values remain valid in such a scenario. According to Lipsitz and Fitzmaurice (2008), the data are said to be MCAR when the probability that responses are missing is unrelated to either the specific values that, in principle, should have been obtained (the missing responses) or the set of observed responses. In Section 2.4.2, the weighted GEE method to handle dropout-type missing data under the MAR (*Missing At Random*) assumption is approached. The MAR assumption is weaker than MCAR.

3.5 Variance estimation

The vcov-type extraction method associated with the objects generated by the function `glmgee()` allows the user to compute five different estimates of $\text{Var}(\hat{\boldsymbol{\beta}})$. The user may specify the estimate type through the argument `type` of the vcov-type method. The five types of estimates for $\text{Var}(\hat{\boldsymbol{\beta}})$ are described as follows:

- `type=“model”`:

$$\hat{\text{V}}\text{ar}_{\text{M}}(\hat{\boldsymbol{\beta}}) = [\mathbf{K}(\hat{\boldsymbol{\beta}})]^{-1} = \hat{\phi}\left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1}$$

- `type=“robust”` (Liang and Zeger, 1986):

$$\hat{\text{V}}\text{ar}_{\text{R}}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n} \mathbf{X}_i^\top \hat{\mathbf{W}}_i \hat{\mathbf{K}}_i^{-1} \mathbf{e}_i \mathbf{e}_i^\top \hat{\mathbf{K}}_i^{-1} \hat{\mathbf{W}}_i \mathbf{X}_i\right)\left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1},$$

  where $\mathbf{e}_i = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_i$. This estimator is robust to misspecification of the working correlation matrix. That is, it is a consistent estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$ provided that the mean model (μ) is correctly specified.

- `type=“df-adjusted”`:

$$\hat{\text{V}}\text{ar}_{\text{A}}(\hat{\boldsymbol{\beta}}) = \frac{n}{n-p-1} \hat{\text{V}}\text{ar}_{\text{R}}(\hat{\boldsymbol{\beta}})$$

- `type=“bias-corrected”` (Mancl and DeRouen, 2001):

$$\hat{\text{V}}\text{ar}_{\text{B}}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n} \mathbf{X}_i^\top \hat{\mathbf{W}}_i \hat{\mathbf{K}}_i^{-1} \widetilde{\mathbf{e}}_i \widetilde{\mathbf{e}}_i^\top \hat{\mathbf{K}}_i^{-1} \hat{\mathbf{W}}_i \mathbf{X}_i\right)\left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1},$$

  where $\widetilde{\mathbf{e}}_i = (\mathbf{I} - \hat{\mathbf{H}}_i)^{-1}\mathbf{e}_i$ and $\hat{\mathbf{H}}_i = \hat{\mathbf{K}}_i \mathbf{X}_i \left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1} \mathbf{X}_i^\top \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1}$. The “bias-corrected” estimator for $\text{Var}(\hat{\boldsymbol{\beta}})$ is also robust to the misspecification of the working correlation matrix, and is very useful in cases where the sample size is “small” due to its improved finite sample properties.

- `type=“jackknife”` (Lipsitz et al., 1990):

$$\hat{\text{V}}\text{ar}_{\text{J}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \hat{\boldsymbol{\beta}}^1_{(i)} - \overline{\hat{\boldsymbol{\beta}}^1}\right)\left(\sum_{i=1}^{n} \hat{\boldsymbol{\beta}}^1_{(i)} - \overline{\hat{\boldsymbol{\beta}}^1}\right)^\top = \hat{\text{V}}\text{ar}_{\text{B}}(\hat{\boldsymbol{\beta}}) - \frac{1}{n}\left(\sum_{i=1}^{n} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(i)}\right)\left(\sum_{i=1}^{n} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(i)}\right)^\top,$$

where $\hat{\boldsymbol{\beta}}^1_{(i)}$ is the "one-step approximation" of $\hat{\boldsymbol{\beta}}_{(i)}$ given in Section 2.3.8, with $\hat{\boldsymbol{\beta}}_{(i)}$ representing the estimate of $\boldsymbol{\beta}$ obtained from the dataset in which the ith cluster or subject is excluded, and $\overline{\hat{\boldsymbol{\beta}}^1} = n^{-1}(\hat{\boldsymbol{\beta}}^1_{(1)} + \ldots + \hat{\boldsymbol{\beta}}^1_{(n)})$.

The summary-type method associated with the objects generated by the function `glmgee()` also allows the user to choose among the five different types of estimates for $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ through its argument `varest`.

3.6 Comparison of nested models

The package `glmtoolbox` includes an anova-type method associated with the objects generated by the function `glmgee()`, which allows the user to compare nested GEE models (that is, it allows the user to assess the hypothesis system $\mathrm{H}_0 : \boldsymbol{\beta}^* = \mathbf{0} \quad \text{versus} \quad \mathrm{H}_1 : \boldsymbol{\beta}^* \neq \mathbf{0}$, where the elements of $\boldsymbol{\beta}^*$ are a subset of those of $\boldsymbol{\beta}$, as $\boldsymbol{\beta}^*$ may be written as $\boldsymbol{\beta}^* = \mathbf{L}^\top\boldsymbol{\beta}$, in which $\mathbf{L}$ is a $r \times (p+1)$ contrast matrix) by using not just the Wald test but also the generalized score test (Rotnitzky and Jewell, 1990; Boos, 1992)). The following decision rule may be used to assess the hypothesis system $\mathrm{H}_0 : \boldsymbol{\beta}^* = \mathbf{0} \quad \text{versus} \quad \mathrm{H}_1 : \boldsymbol{\beta}^* \neq \mathbf{0}$:

"*Reject* H_0 *at the approximate* $100(\alpha)\%$ *significance level if* $\xi > \chi^2_{1-\alpha}(r)$",

where $\alpha \in (0,1)$, $\chi^2_{1-\alpha}(r)$ is the $100(1-\alpha)$th percentile of the chi-square distribution with r degrees-of-freedom, and ξ is one of the following statistics:

- `test="wald"`. Computes the Wald test, which is based on the following statistic:

$$\xi_W = \left(\mathbf{L}^\top\hat{\boldsymbol{\beta}}\right)^\top \left(\mathbf{L}^\top \hat{\mathrm{V}}\mathrm{ar}_{\mathrm{R}}(\hat{\boldsymbol{\beta}})\mathbf{L}\right)^{-1} \left(\mathbf{L}^\top\hat{\boldsymbol{\beta}}\right).$$

- `test="score"`. Computes the generalized score test, whose statistic, denoted here by ξ_S, reduces to the following expression evaluated at the estimate of $\boldsymbol{\beta}$ obtained under the restriction given by H_0 (that is, the estimate of $\boldsymbol{\beta}$ restricted to $\boldsymbol{\beta}^* = \mathbf{0}$):

$$\mathbf{U}^\top(\boldsymbol{\beta})\left[\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{M}}(\hat{\boldsymbol{\beta}})\mathbf{L}\left(\mathbf{L}^\top\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{R}}(\hat{\boldsymbol{\beta}})\mathbf{L}\right)^{-1}\mathbf{L}^\top\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{M}}(\hat{\boldsymbol{\beta}})\right]\mathbf{U}(\boldsymbol{\beta}).$$

The packages `gee` and `geeM` do not include an anova-type method. On the other hand, the anova-type method available in `geepack()` allows the user to compare nested models using just the Wald test.

3.7 Criteria for choosing a working correlation structure

The selection criteria available in `glmtoolbox` are the following:

- Quasi-likelihood under Independence model Criterion (Pan, 2001):

$$\mathrm{QIC} = -2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_{y_{ij}}^{\hat{\mu}_{ij}} \omega_{ij}\frac{(y_{ij}-\mu_{ij})}{\hat{\phi}\,\mathrm{V}(\mu_{ij})}d\mu_{ij} + 2\,\mathrm{trace}\left\{\hat{\phi}^{-1}(\mathbf{X}^\top\hat{\mathbf{K}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{K}}\mathbf{X})\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{R}}(\hat{\boldsymbol{\beta}})\right\}$$

  The expressions for $\int_{y_{ij}}^{\hat{\mu}_{ij}} \omega_{ij}\frac{(y_{ij}-\mu_{ij})}{\mathrm{V}(\mu_{ij})}d\mu_{ij}$ for the variance functions in Table 2.3.1 are listed in Table 9.1 of McCullagh and Nelder (1989).

- Correlation Information Criterion (Hin and Wang, 2009):

$$\mathrm{CIC} = \mathrm{trace}\left\{\hat{\phi}^{-1}(\mathbf{X}^\top\hat{\mathbf{K}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{K}}\mathbf{X})\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{R}}(\hat{\boldsymbol{\beta}})\right\}$$

- Gosho-Hamada-Yoshimura's Criterion (Gosho et al., 2011; Gosho, 2014):

$$\mathrm{GHYC} = \mathrm{trace}\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i-\hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i-\hat{\boldsymbol{\mu}}_i)^\top\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}\hat{\mathbf{V}}_i\right)^{-1} - \mathbf{I}\right]^2\right\}$$

- Pardo-Alonso's Criterion (Pardo and Alonso, 2019):

$$\mathrm{PAC} = \left| \frac{\det\left(\frac{1}{n}\sum\limits_{i=1}^{n}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top\right)}{\det\left(\frac{1}{n}\sum\limits_{i=1}^{n}\hat{\phi}\hat{\mathbf{V}}_i\right)} - 1 \right|$$

- Rotnitzky-Jewell's Criterion (Hin et al., 2007):

$$\mathrm{RJC} = \left(\left[1 - \frac{\mathrm{trace}(\overline{\mathrm{RJC}})}{p+1}\right]^2 + \left[1 - \frac{\mathrm{trace}(\overline{\mathrm{RJC}}^2)}{p+1}\right]^2\right)^{\frac{1}{2}},$$

  where $\overline{\mathrm{RJC}} = \hat{\mathrm{V}}\mathrm{ar}_{\mathrm{R}}(\hat{\boldsymbol{\beta}})[\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{M}}(\hat{\boldsymbol{\beta}})]^{-1}$.

- Akaike-type penalized Gaussian Pseudo-likelihood Criterion (Carey and Wang, 2011; Zhu and Zhu, 2013; Fu et al., 2018):

$$\mathrm{AGPC} = \sum_{i=1}^{n}\left[n_i \log(2\pi) + \frac{1}{\hat{\phi}}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top \hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) + \log(\hat{\phi}|\hat{\mathbf{V}}_i|)\right] + 2(p+1+q)$$

- Schwarz-type penalized Gaussian Pseudo-likelihood Criterion (Carey and Wang, 2011; Zhu and Zhu, 2013; Fu et al., 2018):

$$\mathrm{SGPC} = \sum_{i=1}^{n}\left[n_i \log(2\pi) + \frac{1}{\hat{\phi}}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top \hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) + \log(\hat{\phi}|\hat{\mathbf{V}}_i|)\right] + \log(n)(p+1+q)$$

The above criteria may be computed for one or more GEE models using the extraction methods `QIC()`, `CIC()`, `GHYC()`, `PAC()`, `RJC()`, `AGPC()` and `SGPC()`.

3.8 Global influence

The dfbeta- and cooks.distance-type extraction methods associated with the objects generated by the function `glmgee()` compute and, optionally, display plots of the statistics dfbeta and Cook's distance, which are "leave-one-out" statistics computed to quantify the effect on the estimates of the parameters of interest of deleting each subject/cluster or observation. If the ith cluster is excluded, these statistics may be expressed as follows:

$$\mathrm{Dfbeta}_{(i)} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} \quad \text{and} \quad \mathrm{CD}_{(i)} = \frac{1}{(p+1)}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top [\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}})]^{-1}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}),$$

respectively, where $\hat{\boldsymbol{\beta}}_{(i)}$ represents the estimate of $\boldsymbol{\beta}$ computed from the dataset in which the ith cluster has been excluded. Similar to the extraction function `vcov()`, the estimate of $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ to be used in the computation of the Cook's distance can be chosen by using the argument `varest` of the function `cooks.distance()`, whose options are `varest="model"`, `varest="robust"`, `varest="df-adjusted'`, `varest="bias-corrected'` and `varest="jackknife"`. To avoid computational burden, the values of $\hat{\boldsymbol{\beta}}_{(i)}$ are replaced by their "one-step approximations", denoted here by $\hat{\boldsymbol{\beta}}^1_{(i)}$. Next, the two methods for the computation of the "one-step approximations" available in the dfbeta- and cooks.distance-type extraction methods (through their arguments `method`) are described:

- `method="full"`. $\hat{\boldsymbol{\beta}}_{(i)}$ is replaced by the result of the first iteration of the estimating algorithm of the GEE when it is performed using: (i) the dataset in which the ith cluster has been excluded; and (ii) a starting value which is the solution to the same GEE but computed from the dataset including all clusters (that is, the current $\hat{\boldsymbol{\beta}}$).
- `method="Preisser-Qaqish"`. $\hat{\boldsymbol{\beta}}_{(i)}$ is replaced by the result of the first iteration of the estimating algorithm of the GEE when it is performed using: (i) the dataset in which the ith cluster has been excluded; (ii) a starting value which is the solution to the same GEE but computed from the dataset including all clusters (that is, the current $\hat{\boldsymbol{\beta}}$); and (iii) the working correlation matrix is assumed to be known and equal to its current estimate. According to Preisser and Qaqish (1996); Hammill and Preisser (2006), $\hat{\boldsymbol{\beta}}^1_{(i)}$ reduces to

$$\hat{\boldsymbol{\beta}}^1_{(i)} = \hat{\boldsymbol{\beta}} - \left(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}\right)^{-1} \mathbf{X}_i^\top \hat{\mathbf{W}}_i \hat{\mathbf{K}}_i^{-1} \tilde{\mathbf{e}}_i.$$

Thus, $\hat{\text{V}}\text{ar}_{\text{B}}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} (\text{Dfbeta}_{(i)})(\text{Dfbeta}_{(i)})^\top$. Similar, but more complicated closed-form expressions for $\hat{\boldsymbol{\beta}}^1_{(ij)}$ are given in Preisser and Qaqish (1996) and Hammill and Preisser (2006) when observations instead of clusters/subjects have been excluded from the dataset. The statistics based on those values of $\hat{\boldsymbol{\beta}}_{(ij)}$ may be obtained by specifying `level`=“observations” when using the extraction methods `dfbeta()` and `cooks.distance()`.

3.9 Local influence

The localInfluence-type extraction method associated with the objects generated by the function `glmgee()` computes and, optionally, displays plots of some local influence measures based on the approach proposed by Cook (1986). Let $\mathbf{u}$ be a set of perturbations applied to the model and/or the data. The resulting estimating equations are denoted by $\mathbf{U}(\boldsymbol{\beta}|\mathbf{u})$. Similar to Jung (2008), the following is assumed: (i) the working correlation matrix is known and equal to its current estimate; (ii) the quasi-likelihood function $Q(\boldsymbol{\beta})$ exists such that $\mathbf{U}(\boldsymbol{\beta})$ is its gradient; and (iii) $\mathbf{u}_0$ exists, such that $\mathbf{U}(\boldsymbol{\beta}|\mathbf{u}_0)$ and $\mathbf{U}(\boldsymbol{\beta})$ coincide. So, the influence of the set of perturbations $\mathbf{u}$ on the estimate of $\boldsymbol{\beta}$ may be assessed by studying the conformal normal curvature, $C_{\mathbf{d}}$, around $\mathbf{u}_0$, along a unitary direction $\mathbf{d}$, in which $C_{\mathbf{d}} = 2|\mathbf{d}^\top \mathbf{C}\,\mathbf{d}|$ and

$$\mathbf{C} = \hat{\boldsymbol{\Delta}}^\top[-\ddot{\mathbf{Q}}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\Delta}},$$

where $\hat{\boldsymbol{\Delta}}$ corresponds to the $(p+1) \times \dim(\mathbf{u})$ matrix given by $\dfrac{\partial \mathbf{U}(\boldsymbol{\beta}|\mathbf{u})}{\partial \mathbf{u}^\top} = (\Delta_1, \ldots, \Delta_n)$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\mathbf{u} = \mathbf{u}_0$, and $\ddot{\mathbf{Q}}(\hat{\boldsymbol{\beta}})$ corresponds to the $(p+1) \times (p+1)$ matrix given by $\dfrac{\partial \mathbf{U}(\boldsymbol{\beta}|\mathbf{u})}{\partial \boldsymbol{\beta}^\top}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\mathbf{u} = \mathbf{u}_0$. The matrix $\ddot{\mathbf{Q}}(\boldsymbol{\beta})$ may be written as follows

$$\ddot{\mathbf{Q}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{X}_i^\top \left[\text{diag}\{\mathbf{D}_{(\mathbf{a_i})}\mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\} + \mathbf{K}_i\mathbf{V}_i^{-1}\mathbf{D}_{(\mathbf{b_i})}\right]\mathbf{X}_i$$

in which $\mathbf{D}_{(\mathbf{a_i})}$ and $\mathbf{D}_{(\mathbf{b_i})}$ are diagonal matrices with diagonal elements given by $a_{i1}, \ldots, a_{in_i}$ and $b_{i1}, \ldots, b_{in_i}$, respectively, where

$$a_{ij} = -\frac{1}{[g'(\mu_{ij})]^2}\left[\frac{g''(\mu_{ij})}{g'(\mu_{ij})} + \frac{\text{V}'(\mu_{ij})}{2\text{V}(\mu_{ij})}\right] \quad \text{and} \quad b_{ij} = -\frac{1}{g'(\mu_{ij})}\left[1 + \frac{(y_{ij} - \mu_{ij})\text{V}'(\mu_{ij})}{2\text{V}(\mu_{ij})}\right].$$

Next, the options for the perturbation schemes available in the extraction method `localInfluence()` (through its argument `perturbation`) are described:

- Case weight perturbations

  (1) Clusters (`perturbation=“cw-clusters”`):

  $$\mathbf{U}(\boldsymbol{\beta}|\mathbf{u}) = \sum_{i=1}^{n} u_i \mathbf{X}_i^\top \mathbf{K}_i \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i).$$

  Therefore, $\dim(\mathbf{u}) = n$, $\mathbf{u}_0 = (1, \ldots, 1)^\top$ and $\hat{\Delta}_i = \mathbf{X}_i\hat{\mathbf{K}}_i\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)$ is a $(p+1) \times 1$ matrix.

  (2) Observations (`perturbation=“cw-observations”`)

  $$\mathbf{U}(\boldsymbol{\beta}|\mathbf{u}) = \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{K}_i \text{diag}(\mathbf{u}_i)\mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i),$$

  in which $\mathbf{u_i} = (u_{i1}, \ldots, u_{in_i})^\top$. Therefore, $\dim(\mathbf{u}) = N$, $\mathbf{u_0} = (1, \ldots, 1)^\top$ and $\hat{\Delta}_i = \mathbf{X}_i^\top \text{diag}\{\hat{\mathbf{K}}_i\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)\}$ is a $(p+1) \times n_i$ matrix.

- Response perturbation (`perturbation=“response”`):

  If the response is continuous, then the value of y_{ij} is perturbed by adding a quantity which is proportional to the standard deviation of Y_{ij}, thus,

  $$\mathbf{U}(\boldsymbol{\beta}|\mathbf{u}) = \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{K}_i \mathbf{V}_i^{-1}(\mathbf{y}_i + \sqrt{\phi}[\text{diag}(\mathbf{V}_i)]^{\frac{1}{2}}\mathbf{u}_i - \boldsymbol{\mu}_i),$$

  in which $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$. Therefore, $\dim(\mathbf{u}) = N$, $\mathbf{u}_0 = (0, \ldots, 0)^\top$ and $\hat{\Delta}_i = \hat{\phi}^{\frac{1}{2}}\mathbf{X}_i^\top \hat{\mathbf{K}}_i\hat{\mathbf{V}}_i^{-1}[\text{diag}(\hat{\mathbf{V}}_{\text{i}})]^{\frac{1}{2}}$ is a $(p+1) \times n_i$ matrix.

- Perturbation of covariates (`perturbation= “covariate”`):
  If the r-column of $\mathbf{X}$ corresponds to a continuous covariate, then the values of that covariate are perturbed by adding a quantity which is proportional to its standard deviation, c_r, thus,

  $$\mathbf{U}(\boldsymbol{\beta}|\mathbf{u}) = \sum_{i=1}^{n} \left(\mathbf{X}_i + \mathbf{u}_i\delta_r^\top\right) \mathbf{K}_i\mathbf{V}_i^{-1} \left(\mathbf{y}_i - g^{-1}\left[\mathbf{z}_i + (\mathbf{X}_i + \mathbf{u}_i\delta_r^\top)\boldsymbol{\beta}\right]\right),$$

  in which $\mathbf{u}_i = (u_{i1},\ldots,u_{in_i})^\top$ and δ_r is a $(p+1)$-dimensional column vector of zeros with the known constant c_r in the r-th position. Therefore, $\dim(\mathbf{u}) = N$, $\mathbf{u}_0 = (0,\ldots,0)^\top$ and $\Delta_i = c_r\delta_r\left\{(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{K}}_i + (\mathbf{X}_i\hat{\boldsymbol{\beta}})^\top\left[\hat{\mathbf{M}}_i\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) - \hat{\mathbf{K}}_i\hat{\mathbf{V}}_i^{-1}\hat{\mathbf{K}}_i\mathbf{1}_{n_i}\right]\mathbf{1}_{n_i}^\top\right\}$ is a $(p+1)\times n_i$ matrix, where $\mathbf{M}_i$ is a diagonal matrix with diagonal elements $m_{i1},\ldots,m_{in_i}$, with $m_{ij} = -g''(\mu_{ij})/[g'(\mu_{ij})]^3$.

The options to study the local influence from $C_{\mathbf{d}}$ available in the extraction method `localInfluence()` (through its argument `type`) are described as follows:

- `type=“local”`. Computes and, optionally, displays an index plot of the elements of $\mathbf{d}_{\max}$ (that is, the eigenvector which correspond to the maximum absolute eigenvalue of $\mathbf{C}$). The vector $\mathbf{d}_{\max}$ is computed using the power iteration algorithm.
- `type=“total”`. Computes and, optionally, displays an index plot of the elements of the main diagonal of $\mathbf{C}$.

3.10 Variable selection

The function `stepCriterion()` associated with the objects generated by the function `glmgee()` allows the user to iteratively choose the more “relevant” and/or ” significant” variables and/or effects in the model fit by using either of the following “hybrid stepwise” strategies (see James et al. (2013), page 212):

- `direction=“forward”`. The “hybrid forward stepwise” strategy starts with the simplest model (which may be set at the argument `scope`, and by default, is a model whose parameters in the linear predictor, except the intercept, if there is, are set to be 0), and then the candidate models are built by hierarchically including effects in the linear predictor, whose “relevance” and/or “importance” in the model fit is assessed by comparing nested models (that is, by comparing the models with and without the included effect) using a criterion previously specified. If an effect is included in the model, this strategy may also remove any effect which, according to the previously specified criterion, no longer contributes to an improvement in the model fit.
- `direction=“backward”`. The “hybrid backward stepwise” strategy starts with the more complex model (which may be specified at the argument `scope`), and then the candidate models are built by hierarchically removing effects in the linear predictor, whose “relevance” and/or “importance” in the model fit is assessed by comparing nested models (that is, by comparing the models with and without the excluded effect) using a criterion previously chosen. If an effect is excluded from the model, then this strategy may also add any effect which, according to the criterion previously specified, provides an improvement in the model fit.

The available comparison criteria are the following:

- `criterion=“qic”`. QIC
- `criterion=“qicu”`. According to Pan (2001), the QICu may be written as

  $$\text{QICu} = -2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_{y_{ij}}^{\hat{\mu}_{ij}} \omega_{ij}\frac{(y_{ij}-\mu_{ij})}{\phi\,\mathrm{V}(\mu_{ij})}d\mu_{ij} + 2(p+1)$$

- `criterion=“agpc”`. AGPC
- `criterion=“sgpc”`. SGPC
- `criterion=“p-value”`. p-values of the Wald (`test=“wald”`) or generalized score (`test=“score”`) tests.

According to Xu et al. (2019), the AGPC and SGPC outperform other existing methods in selecting variables, and they perform well regardless of whether the working correlation structure is correctly specified or not.

3.11 Leverage

The leverage-type extraction method associated with the objects generated by the function `glmgee()` computes and, optionally, displays a plot of the leverage measures at the cluster- and observation-level. According to Preisser and Qaqish (1996); Hammill and Preisser (2006), the observation leverage for the jth observation of the ith cluster is the value of the jth diagonal element of the matrix $\hat{\mathbf{H}}_i$. Cluster leverage for cluster i is the mean of the observation leverages. This, unlike the sum of observation leverages, makes the leverage measures comparable when there are clusters of different sizes. The leverage at cluster- and observation-level may be obtained from the leverage-type extraction method (`leverage()`) by specifying `level="clusters"` and `level="observations"`, respectively.

3.12 Residuals

The residuals-type extraction method associated with the objects generated by the function `glmgee()` computes and, optionally, displays a plot of three different types of residuals. The user may specify the residual type through the argument `type` of the residuals-type method. The residuals are computed to quantify the goodness-of-fit of the model at the cluster-level (Mahalanobis-type residuals) and at the observation-level (Pearson- and deviance-type residuals). Indeed, a plot of the Pearson- or deviance-type residuals versus the fitted values may be a useful tool to assess if, for instance, the specified variance function provides a suitable description of the dispersion present in the data. As follows, three types of residuals are described:

- `type="pearson"`. Computes the Pearson-type residuals, given by

$$r_{ij}^{P} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\phi}\,\mathrm{V}(\hat{\mu}_{ij})/\omega_{ij}}} \qquad \text{for} \quad i = 1,\ldots,n \quad \text{and} \quad j = 1,\ldots,n_i.$$

- `type="deviance"`. Computes the deviance-type residuals, given by

$$r_{ij}^{D} = \mathrm{sign}(y_{ij} - \hat{\mu}_{ij})\sqrt{d(y_{ij},\hat{\mu}_{ij},\omega_{ij})/\hat{\phi}} \qquad \text{for} \quad i = 1,\ldots,n \quad \text{and} \quad j = 1,\ldots,n_i,$$

  where $d(y_{ij},\mu_{ij},\omega_{ij})$ represents the contribution to the non-scaled deviance of the jth measurement performed on the ith subject or cluster.

- `type="mahalanobis"`. Computes the Mahalanobis-type residuals, given by

$$r_i^{M} = n_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^{\top}[\hat{\mathrm{V}}\mathrm{ar}(Y_i)]^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) = \frac{n_i^{-1}}{\hat{\phi}}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^{\top}\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) \qquad \text{for} \quad i = 1,\ldots,n.$$

The residuals-type extraction method in `geepack` provides neither deviance- nor Mahalanobis-type residuals computation, whereas `gee` and `geeM` do not include a residuals-type extraction method.

4 Extensions

4.1 Nonlinear predictors

Unlike that described in expression (1), where $g(\mu_{ij})$ is restricted to being a linear combination of the elements of the parameter vector $\boldsymbol{\beta}$, in the new model formulation described here, $g(\mu_{ij})$ may be expressed using a more general family of functions of $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^{\top}$. That is,

$$g(\mu_{ij}) = \eta(\mathbf{x}_{ij},\boldsymbol{\beta}),$$

where $\eta_{ij}(\boldsymbol{\beta}) \equiv \eta(\mathbf{x}_{ij},\boldsymbol{\beta})$ is a continuous and twice differentiable function of $\boldsymbol{\beta}$, with $\mathbf{x}_{ij}$ being a vector of continuous and/or discrete regressors. The estimate of $\boldsymbol{\beta}$ can be obtained by solving the p non-linear equations given by $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, where

$$\mathbf{U}(\boldsymbol{\beta}) = \phi^{-1}\sum_{i=1}^{n}\mathbf{D}_i^{\top}\mathbf{K}_i\mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \phi^{-1}\sum_{i=1}^{n}\mathbf{D}_i^{\top}\mathbf{W}_i\mathbf{K}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \phi^{-1}\mathbf{D}^{\top}\mathbf{W}\mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

in which $\mathbf{D} = (\mathbf{D}_1^{\top},\ldots,\mathbf{D}_n^{\top})^{\top}$, $\mathbf{D}_i = (\mathbf{d}_{i1},\ldots,\mathbf{d}_{in_i})^{\top}$ and $\mathbf{d}_{ij} = (\partial\eta_{ij}(\boldsymbol{\beta})/\partial\beta_1,\ldots,\partial\eta_{ij}(\boldsymbol{\beta})/\partial\beta_p)^{\top}$. This type of GEE model is implemented in the function `gnmgee()` of the package `glmtoolbox`. The arguments of the function `gnmgee()` are very similar to those of `glmgee()`. Nevertheless, the form of the non-linear function $\eta_{ij}(\boldsymbol{\beta})$ and the starting value for $\boldsymbol{\beta}$ in the estimation algorithm must be set by the user via the arguments `formula` and `start` of `gnmgee()`. But, the argument `formula` also accepts built-in non-linear functions such as `SSasymp()`, `SSasympOff()`, `SSasympOrig()`, `SSbiexp()`, `SSfol()`, `SSfpl()`, `SSgompertz()`, `SSlogis()`, `SSmicmen()` and `SSweibull()`, which do not require user-supplied starting values.

4.2 Weighted GEE methods

In longitudinal studies, in which the response variable is planned to be measured at J time points on each subject/cluster, the weighted GEE methods provide consistent estimates under the MAR assumption when the missing data pattern is dropout and its mechanism is correctly specified (Robins et al., 1995). Let $t_{ij} = 1$ if the response is observed on the ith subject/cluster at time j, and 0 otherwise, which is assumed to be a realization of a random variable denoted here by T_{ij}. In addition, let $t_i = 1 + t_{i1} + t_{i2} + \ldots + t_{iJ}$ the time of dropout for the ith subject/cluster, so, $t_i \in \{2, \ldots, J+1\}$. The probability of observing y_{ij} may be expressed using the following logistic model:

$$\text{logit}(\pi_{ij}) = \text{logit}\Big(\Pr\Big[T_{ij} = 1 \mid T_{i,j-1} = 1, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij}, Y_{i1}, \ldots, Y_{i,j-1}\Big]\Big) = \mathbf{z}_{ij}^\top \boldsymbol{\tau},$$

where $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_s)^\top$ is an unknown parameter vector and $\mathbf{z}_{ij} = (1, z_{ij1}, \ldots, z_{ijs})$ is a vector of regressors which may include visit indicator variables ($I(j = 2)$,...,$I(j = J-1)$, where $I(\cdot)$ is the indicator function), covariates $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij})$ and past responses $(y_{i1}, \ldots, y_{i,j-1})$. The maximum likelihood estimate of $\boldsymbol{\tau}$ is $\hat{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau} \in \mathbb{R}^{s+1}}{\text{argmax}}\, \ell(\boldsymbol{\tau})$, where (Robins et al., 1995; Preisser et al., 2002)

$$\ell(\boldsymbol{\tau}) = \sum_{i=1}^{n} \sum_{j=1}^{t_i^*} t_{i,j-1} \Big[t_{ij} \log(\pi_{ij}) + (1 - t_{ij}) \log(1 - \pi_{ij})\Big]$$

is the log-likelihood function of $\boldsymbol{\tau}$, in which $t_i^* = \min(t_i, J)$ and $t_{i0} = 0$. If the missing data pattern is dropout, then $t_{i1} = 1$ is assumed for all $i = 1, \ldots .n$. Therefore,

$$\hat{\pi}_{ij} = \begin{cases} 1 & \text{if } j = 1 \\ \dfrac{\exp(\mathbf{z}_{ij}^\top \hat{\boldsymbol{\tau}})}{1 + \exp(\mathbf{z}_{ij}^\top \hat{\boldsymbol{\tau}})} & \text{if } j > 1 \end{cases}$$

The estimate of $\boldsymbol{\tau}$ satisfies $\left.\dfrac{\partial \ell(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}} = \mathbf{S}_1 + \ldots + \mathbf{S}_n = \mathbf{0}$, where $\mathbf{S}_i = \mathbf{Z}_i^\top(\mathbf{t} - \hat{\boldsymbol{\pi}})$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{it_i^*})^\top$, $\mathbf{t} = (t_{i1}, \ldots, t_{it_i^*})^\top$ and $\hat{\boldsymbol{\pi}} = (\hat{\pi}_{i1}, \ldots, \hat{\pi}_{it_i^*})^\top$.

Observation-specific weights

According to Fitzmaurice et al. (2011), the underlying idea of this weighting method is to base the estimation on the observed responses but weight them to account for the probability of remaining in the study. When using the observation-specific weighted GEE method, the covariates for all occasions for a subject/cluster must be observed, regardless of whether the response is missing. That is, the input data set must contain J observations for each subject/cluster. The estimate of $\boldsymbol{\beta}$ is the solution to the $(p+1)$ nonlinear equations given by $\mathbf{U}^*(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, in which $\mathbf{U}^*(\boldsymbol{\beta})$ may be expressed as follows

$$\mathbf{U}^*(\boldsymbol{\beta}) = \phi^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{K}_i \mathbf{V}_i^{-1} \boldsymbol{\Lambda}_i (\mathbf{y}_i - \boldsymbol{\mu}_i),$$

where $\boldsymbol{\Lambda}_i = \text{diag}\{t_{i1}\lambda_{i1}, \ldots, t_{iJ}\lambda_{iJ}\}$ and $\lambda_{ij} = 1/(\hat{\pi}_{i1} \times \hat{\pi}_{i2} \times \ldots \times \hat{\pi}_{ij})$. The estimator of $\boldsymbol{\beta}$ has an asymptotic normal distribution with consistent estimator of its asymptotic variance given by (Robins et al., 1995; Preisser et al., 2002)

$$\hat{\text{V}}\text{ar}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^\top \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \boldsymbol{\Lambda}_i \hat{\mathbf{K}}_i \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{E}_i \mathbf{E}_i^\top\right) \left(\sum_{i=1}^{n} \mathbf{X}_i^\top \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \boldsymbol{\Lambda}_i \hat{\mathbf{K}}_i \mathbf{X}_i\right)^{-1},$$

where $\mathbf{E}_i = \mathbf{X}_i^\top \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \boldsymbol{\Lambda}_i (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) - \left(\sum\limits_{i=1}^{n} \mathbf{X}_i^\top \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \boldsymbol{\Lambda}_i (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) \mathbf{S}_i^\top\right) \left(\sum\limits_{i=1}^{n} \mathbf{S}_i \mathbf{S}_i^\top\right)^{-1} \mathbf{S}_i$.

Subject/Cluster-specific weights

In the cluster-specific weighted GEE method, covariates for a cluster who dropout at time k must be observed for occasions up to and including time k. That is, each subject must have at least k observations in the input data set. The estimate of $\boldsymbol{\beta}$ is the solution to the $(p+1)$ nonlinear equations given by $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, in which $\mathbf{U}(\boldsymbol{\beta})$ is given by (2), but where the weights in the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are set to be $\omega_{ij}^{*} = \omega_{ij} \times \lambda_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, t_i - 1$. The value of λ_i may be computed as follows

$$\lambda_i^{-1} = \left(\prod_{j=1}^{t_i-1} \hat{\pi}_{ij}\right)(1 - \hat{\pi}_{it_i})^{I(t_i \leq J)}.$$

The estimator of $\boldsymbol{\beta}$ has an asymptotic normal distribution with a consistent estimator of its asymptotic variance given by (Robins et al., 1995; Preisser et al., 2002)

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^{\top} \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{K}}_i \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{E}_i \mathbf{E}_i^{\top}\right) \left(\sum_{i=1}^{n} \mathbf{X}_i^{\top} \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{K}}_i \mathbf{X}_i\right)^{-1},$$

where $\mathbf{E}_i = \mathbf{X}_i^{\top} \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) - \left(\sum\limits_{i=1}^{n} \mathbf{X}_i^{\top} \hat{\mathbf{K}}_i \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) \mathbf{S}_i^{\top}\right) \left(\sum\limits_{i=1}^{n} \mathbf{S}_i \mathbf{S}_i^{\top}\right)^{-1} \mathbf{S}_i.$

The weighted GEE methods based on observation-specific weights and cluster-specific weights are implemented in the function `wglmgee()` of the package `glmtoolbox`, whose arguments are very similar to those of the function `glmgee()`. In that function, the user sets the GEE and missingness models at the same argument of type `Formula` (Zeileis and Croissant, 2010). In addition, the user sets the type of weighting method: observation-specific weights (`level=`“`observations`”) or cluster-specific weights (`level=`“`clusters`”). The `wglmgee()` function estimates the parameters of the missingness model and uses them to compute the required weights. Then, `wglmgee()` introduces the weights in the estimation process of the GEE model parameters as well as in the estimation of their asymptotic variance matrix.

5 Examples

5.1 Growth patterns of trees under two types of atmosphere

The dataset of this example, described by Diggle et al. (2002) and available in the `spruces` object of `glmtoolbox`, is composed of the columns `tree`, `days`, `size` and `treat` (see Table 1). The analysis of this dataset aims to assess the effect of ozone pollution on tree growth. Ozone pollution is common in urban areas, thus the impact of increased ozone concentrations on tree growth is of considerable interest. The response variable is tree size (`size`), where size is conventionally measured by the product of tree height and stem diameter squared. A total of 79 trees, identified in the dataset by the column `tree`, were considered in this experiment. In the first group a total of 54 trees were grown under an ozone-enriched atmosphere (`treat=`“ozone-enriched”), that is, ozone exposure at 70 parts per billion, whereas in the second group, a total of 25 trees were grown under a normal atmosphere (`treat=`“normal”). The size of each tree was observed and recorded exactly 13 times across the time since the experiment began (`days`), so the data are balanced and the number of rows in the dataset is 1027. The main objective of the analysis is to compare the growth patterns of trees under two types of atmosphere: normal and ozone-enriched.

Column	Role	Description
`tree`	Cluster/subject identifier	Identifier of the tree
`days`	Explanatory variable	Number of days after the treatment began
`treat`	Explanatory variable	Treatment: “normal” or “ozone-enriched”
`size`	Response variable	Tree size

Table 1: Columns in the object `spruces` of the package `glmtoolbox`.

An adjusted for skewness box-plot of the data (Figure 1), obtained using the function `adjbox()` in the package `robustbase` (Maechler et al. (2022)), shows that ozone suppresses tree growth. The plot also indicates that under the two types of atmosphere the location as well as the dispersion of the tree size are non-decreasing and non-linear functions of the time since the experiment began, which suggests the data should be analyzed using a GEE with the following features: (i) a linear predictor which includes a polynomial effect of the time, as well as a dummy variable to indicate the type of atmosphere under which the trees grew; and (ii) an increasing variance function such as $\mathrm{V}(\mu) = \mu$,

$V(\mu) = \mu^2$ or $V(\mu) = \mu^3$, which is aimed to include in the model the heteroscedasticity observed in the data.

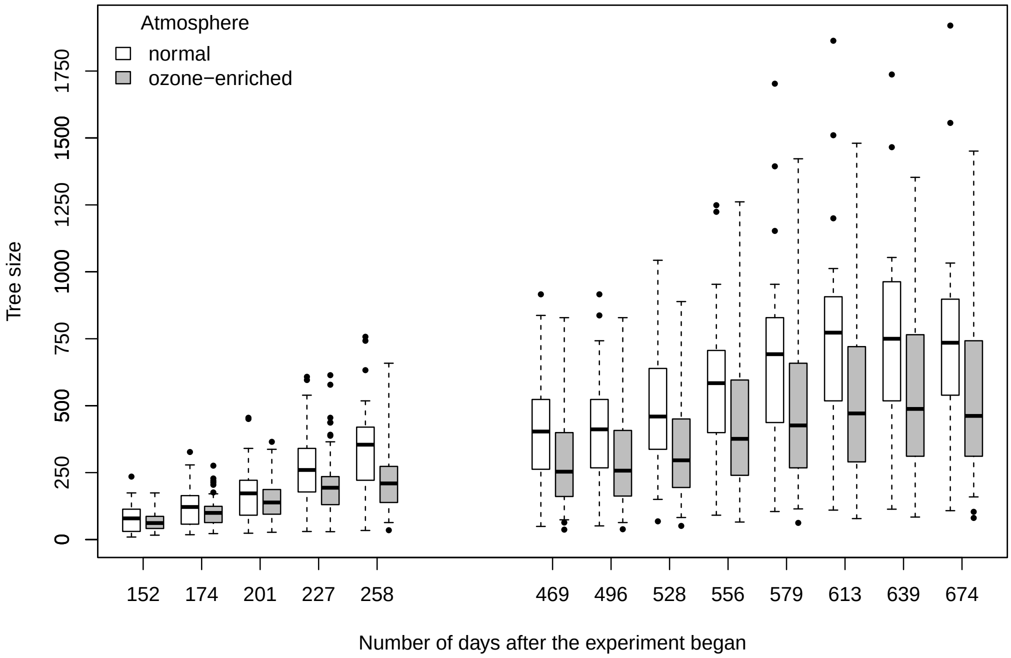

Figure 1: Growth patterns of trees under normal and ozone-enriched atmospheres.

Plots (not shown here) of the deviance-type residuals versus the fitted values for two GEE with logarithmic link, working correlation matrix specified to be the identity matrix, and variance functions $V(\mu) = \mu$ and $V(\mu) = \mu^3$, reveal megaphone shaped and inverted megaphone shaped patterns, respectively, indicating thus how inappropriate such variance functions are for describing the heteroscedasticity present in the data. On the other hand, the same plot but using the variance function $V(\mu) = \mu^2$ (Figure 2(a)) does not reveal any trend, indicating that the data could be suitably analyzed under the assumption of a constant coefficient of variation. So, several GEE with the variance function $V(\mu) = \mu^2$, logarithmic link, and different structures for the working correlation matrix are fitted to the data. Then, the selection criteria available in `glmtoolbox` are used to choose the more suitable structure for the correlation matrix.

```
> data(spruces)
> m1 <- glmgee(size ~ poly(days,4) + treat, id=tree, family=Gamma(log), data=spruces)
> m2 <- update(m1, corstr="Exchangeable")
> m3 <- update(m1, corstr="AR-M-dependent(1)")
> m4 <- update(m1, corstr="AR-M-dependent(2)")
> m5 <- update(m1, corstr="AR-M-dependent(3)")

> a <- CIC(m1, m2, m3, m4, m5, verbose=FALSE)
> b <- QIC(m1, m2, m3, m4, m5, verbose=FALSE)
> c <- GHYC(m1, m2, m3, m4, m5, verbose=FALSE)
> d <- RJC(m1, m2, m3, m4, m5, verbose=FALSE)
> e <- AGPC(m1, m2, m3, m4, m5, verbose=FALSE)
> f <- SGPC(m1, m2, m3, m4, m5, verbose=FALSE)
> cbind(a,QIC=b[,"QIC"],GHYC=c[,"GHYC"],RJC=d[,"RJC"],AGPC=e[,"AGPC"],SGPC=f[,"SGPC"])

  Object       Correlation   CIC   QIC   GHYC    RJC  AGPC  SGPC
1     m1      Independence 23.43 42068 116.42 41.303 13539 13554
2     m2      Exchangeable 23.43 42068  40.96  7.639 11689 11706
3     m3 AR-M-dependent(1) 23.66 42086  11.26  0.129 10941 10957
4     m4 AR-M-dependent(2) 23.56 42158  13.72  0.489 10981 11000
```

```
5     m5 AR-M-dependent(3) 23.56 42201  12.45  0.914 10994 11016
```

Most of the selection criteria (that is, Gosho-Hamada-Yoshimura's criterion, Rotnitzky-Jewell's criterion, Akaike-type penalized Gaussian pseudo-likelihood criterion, and Schwarz-type penalized Gaussian pseudo-likelihood criterion) suggest the first-order autoregressive (AR-1) and Independence as the more and less adequate structures for the correlation matrix, respectively. GEE with the AR-1 structure for the correlation matrix is summarized as follows:

```
> summary(m3)
Sample size
   Number of observations:  1027
       Number of clusters:  79
             Cluster size:  13
*********************************************************************
Model
        Variance function:  Gamma
            Link function:  log
    Correlation structure:  AR-M-dependent(1)
*********************************************************************
Coefficients
                     Estimate Std.Error   z-value Pr(>|z|)
(Intercept)           5.90378   0.10486  56.30321  < 2e-16
poly(days, 4)1       19.20015   0.51848  37.03159  < 2e-16
poly(days, 4)2       -2.85755   0.20585 -13.88147  < 2e-16
poly(days, 4)3        5.41639   0.18246  29.68549  < 2e-16
poly(days, 4)4       -3.57407   0.12478 -28.64405  < 2e-16
treatozone-enriched  -0.25861   0.12835  -2.01486 0.043919

Dispersion            0.32866
*********************************************************************
Working correlation
     [1]  [2]  [3]  [4]  [5]  [6]  [7]  [8]  [9]  [10] [11] [12] [13]
[1]  1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78 0.76 0.73 0.70 0.68 0.66
[2]  0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78 0.76 0.73 0.70 0.68
[3]  0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78 0.76 0.73 0.70
[4]  0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78 0.76 0.73
[5]  0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78 0.76
[6]  0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81 0.78
[7]  0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84 0.81
[8]  0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87 0.84
[9]  0.76 0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90 0.87
[10] 0.73 0.76 0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93 0.90
[11] 0.70 0.73 0.76 0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97 0.93
[12] 0.68 0.70 0.73 0.76 0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00 0.97
[13] 0.66 0.68 0.70 0.73 0.76 0.78 0.81 0.84 0.87 0.90 0.93 0.97 1.00
```

The Wald test and the generalized score test suggest that there is no interaction between time and the type of atmosphere. This is because, as shown below, the *p*-values associated with that effect are "large".

```
> m3a <- update(m3, . ~ . + poly(days,4):treat)
> anova(m3a, test="wald")
Wald test

Model 1 :  size ~ 1
Model 2 :  size ~ poly(days, 4)
Model 3 :  size ~ poly(days, 4) + treat
Model 4 :  size ~ poly(days, 4) + treat + poly(days, 4):treat

           Chi    df   Pr(>Chi)
1 vs 2 1931.9813   4    < 2e-16 ***
2 vs 3    4.0597   1    0.04392 *
3 vs 4    3.6641   4    0.45336

> anova(m3a, test="score")
```

```
Generalized score test

Model 1 :  size ~ 1
Model 2 :  size ~ poly(days, 4)
Model 3 :  size ~ poly(days, 4) + treat
Model 4 :  size ~ poly(days, 4) + treat + poly(days, 4):treat

         Chi    df   Pr(>Chi)
1 vs 2 61.3028   4  1.544e-12 ***
2 vs 3  3.3687   1    0.06645 .
3 vs 4  3.4665   4    0.48300
```

Variance estimation

Next, the estimates of $\text{Var}(\hat{\beta}_0), \text{Var}(\hat{\beta}_1), \ldots, \text{Var}(\hat{\beta}_5)$ are obtained using four different estimators.

```
> cbind(model=diag(vcov(m3, type="model")),
+       robust=diag(vcov(m3, type="robust")),
+       bias.corrected=diag(vcov(m3, type="bias-corrected")),
+       jackknife=diag(vcov(m3, type="jackknife")))
                     model robust bias.corrected jackknife
(Intercept)         0.0110 0.0110         0.0119    0.0119
poly(days, 4)1      0.2564 0.2688         0.2758    0.2758
poly(days, 4)2      0.0922 0.0424         0.0435    0.0435
poly(days, 4)3      0.0352 0.0333         0.0342    0.0342
poly(days, 4)4      0.0283 0.0156         0.0160    0.0160
treatozone-enriched 0.0159 0.0165         0.0176    0.0176
```

Parameter interpretation

Across time, the expected size of the trees that grew under the ozone-enriched atmosphere is approximately $22.79\% = 100 \times [1 - \exp(\hat{\beta}^{\text{treat}})]$ lower than that of the trees that grew under the normal atmosphere, where β^{treat} represents the parameter associated with the dummy variable indicating the type of atmosphere under which the trees grew. According to the Wald test, the hypothesis $\text{H}_0 : \beta^{\text{treat}} \geq 0$ versus $\text{H}_1 : \beta^{\text{treat}} < 0$ is rejected at the approximate level of 5%, indicating thus that the ozone-enriched atmosphere suppresses tree growth.

Variable selection

As an illustration, the procedure of variable selection is applied using the strategy "hybrid forward stepwise" with the p-value of the generalized score test as the comparison criterion, and where the thresholds for add and drop effects are set at 10% and 5%, respectively. In addition, the simplest and most complex models are specified to be 1 and `1+poly(days,4)+treat+poly(days,4):treat`, respectively. As shown below, the best linear predictor according to the chosen strategy incorporates both time and atmosphere, but not the interaction between them. The same results are obtained when the strategy of variable selection is changed as follows: (i) the generalized score test is replaced by the Wald test; (ii) and the "hybrid forward stepwise" procedure is replaced by the "hybrid backward stepwise".

```
stepCriterion(m3, direction="forward", criterion="p-value", test="score",
+              scope=list(lower=~1, upper=~poly(days,4)*treat), levels=c(0.10,0.05))

    Variance function:  Gamma
        Link function:  log
Correlation structure:  AR-M-dependent(1)
  Comparison criteria:  P(Chisq>)(*)

Initial model:
~ 1

Step 0 :
                        df   QIC  QICu  AGPC  SGPC P(Chisq>)(*)
+ poly(days, 4)          4 39944 39928 10927 10941    1.544e-12
```

```
+ treat                1 19534 19517 12353 12360      0.03618
<none>                   18998 18989 12362 12367

Step 1 : + poly(days, 4)

                      df   QIC  QICu  AGPC  SGPC P(Chisq>)(*)
+ treat                1 42086 42051 10941 10957      0.06645
<none>                   39944 39928 10927 10941

Step 2 : + treat

                      df   QIC  QICu  AGPC  SGPC P(Chisq>)(*)
+ poly(days, 4):treat  4 41815 41787 10954 10980        0.483
<none>                   42086 42051 10941 10957

Final model:
~ poly(days, 4) + treat
****************************************************************************
(*) p-values of the generalized score test
 Effects are added when their p-values are lower than 0.1
 Effects are excluded when their p-values are higher than 0.05
```

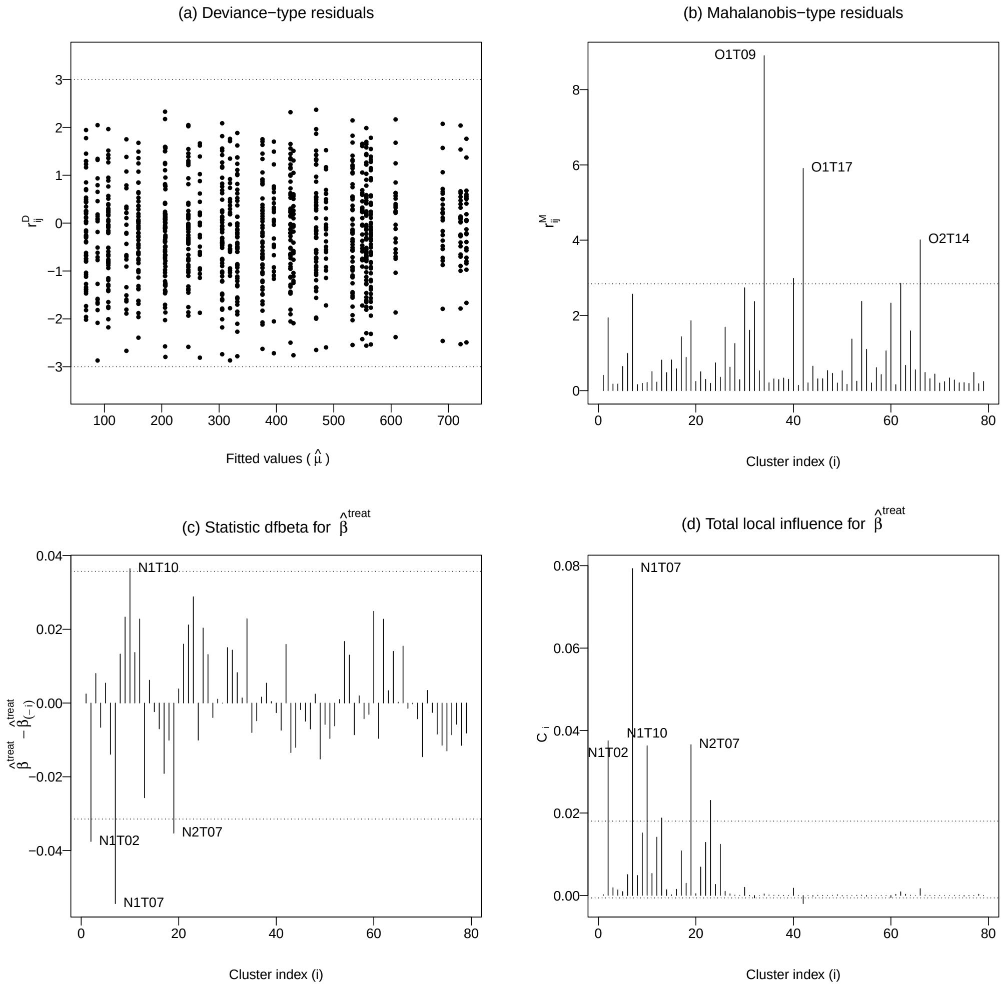

Figure 2: Some diagnostic plots for the GEE with correlation structure AR-1 fitted to the data on growth patterns of trees.

Residual analysis

```
> residuals(m3, type="deviance", plot.it=TRUE, pch=16)
> residuals(m3, type="mahalanobis", plot.it=TRUE, identify=3)
```

According to the Mahalanobis-type residuals (Figure 2(b)) the trees for which the model has the lowest goodness-of-fit are those identified as O1T09, O1T17 and O2T14. Although those trees grew under an ozone-enriched atmosphere, which seems to be associated with expected sizes lower than those of the trees which grew under the normal atmosphere, their observed sizes across the time are even greater than those observed for 70% of the trees which grew under a normal atmosphere.

Global influence

```
> dfbeta(m3, method="full", coefs="treat", identify=4)
```

According to the statistic Dfbeta at the cluster-level for $\hat{\beta}^{\text{treat}}$ (Figure 2(c)), the trees known as N1T02, N1T07 and N2T07 are those providing the greatest evidence which supports the negative effect of the ozone-enriched atmosphere on the growth pattern of the trees, as their exclusion from the dataset leads the estimate of β^{treat} closer to zero. Those trees grew under a normal atmosphere, and their sizes across time are higher than those observed for 90% of the trees growing under such an atmosphere. On the other hand, the tree identified as N1T10 is that providing the greatest evidence against the negative effect of the ozone-enriched atmosphere on the growth pattern of the trees, because its exclusion from the dataset decreases the estimate of β^{treat}. The tree identified as N1T10 grew under a normal atmosphere, however, its size across time is lower than that observed for 90% of the trees growing under the ozone-enriched atmosphere.

Local influence

```
> localInfluence(m3, type="total", perturbation="cw-clusters", coefs="treat",
+                plot.it=TRUE, identify=4)
```

The plot of the total local influence under the case weight perturbation scheme at the cluster-level for $\hat{\beta}^{\text{treat}}$ (Figure 2(d)) highlights the trees identified as N1T07, N1T02, N2T07 and N1T10 as suspected to be influential on the estimate of β^{treat}, which confirms the results of the global influence analysis above.

5.2 Comparison with other GEE solvers

The parameter estimates and the associated standard errors provided by the function `glmgee()` are compared with those generated by the GEE solvers available in the packages `gee`, `geepack` and `geeM`. The results are presented in Table 2. The values provided by the other GEE solvers are very similar to those generated by the function `glmgee()`.

	`glmtoolbox`	`geepack`	`gee`	`geeM`
(Intercept)	5.904(0.105)	5.903(0.105)	5.904(0.105)	5.904(0.105)
poly(days, 4)1	19.200(0.518)	19.186(0.519)	19.201(0.518)	19.200(0.518)
poly(days, 4)2	-2.858(0.206)	-2.860(0.206)	-2.857(0.206)	-2.858(0.206)
poly(days, 4)3	5.416(0.182)	5.414(0.182)	5.417(0.182)	5.416(0.182)
poly(days, 4)4	-3.574(0.125)	-3.572(0.125)	-3.574(0.125)	-3.574(0.125)
treatozone-enriched	-0.259(0.128)	-0.257(0.128)	-0.259(0.128)	-0.259(0.128)
ρ	0.97	0.97	0.97	0.97

Table 2: Parameter estimates (standard errors) of the GEE model with correlation structure AR-1 fitted to the data on growth patterns of trees.

5.3 Treatment of severe postnatal depression

The dataset of this example, available in the object `depression` of `glmtoolbox` and composed of columns named `subj`, `group`, `visit`, `dep` and `depressd` (see Table 3), arose from a double-blind placebo-controlled study on the efficacy of oestrogen given transdermally for treatment of severe postnatal depression (Gregoire et al., 1996). A total of 61 women with severe depression (identified in the dataset by the column `subj`), which began within 3 months of childbirth and persisted for up to 18 months postnatal, were randomly assigned to either a control group (`group="placebo"`) of size 27, which were treated with a placebo patch, or an active treatment group (`group="oestrogen"`) of size 34, which were treated with an oestrogen patch. Prior to therapy, all women were assessed by self-ratings of depressive symptoms on the Edinburgh Postnatal Depression Scale (EPDS), where higher scores are indicative of higher levels of depression. A monthly EPDS (dep) was collected for six months once treatment began (`visit`). The binary response (`depressd`) was 1 to indicate severe depression when the EDPS value was greater than or equal to 11, and 0 in other cases. There are missing values in the data, because for some women, just two measurements of the response variable are available. However, those missing values are positioned at the last time positions, so, there are no intermixed missing values, and the argument `waves` of the function `glmgee()` is not needed.

Column	Role	Description
`subj`	Cluster/subject identifier	Identifier of the woman
`group`	Explanatory variable	Treatment: "placebo" or "oestrogen"
`visit`	Explanatory variable	Number of months after the treatment began
`depressd`	Response variable	1 if EDPS $\geq$ 11 and 0 otherwise

Table 3: Columns in the object `depression` of the package `glmtoolbox`.

A plot of the data (Figure 3) suggests that oestrogen patches are an effective treatment for postnatal severe depression, as across time, the proportion of women with severe depression is lower in the group treated with oestrogen patches than in the group treated with placebo patches. The plot also indicates that the (logit of the) proportion of women with severe depression decreases linearly as a function of the time since the therapy began, but the rate of decreasing seems to be independent of the type of patch (placebo or oestrogen), which suggests that the data could be analyzed by using a GEE with the logit link function and a linear predictor including the effects of time and type of patch, but without the interaction between them.

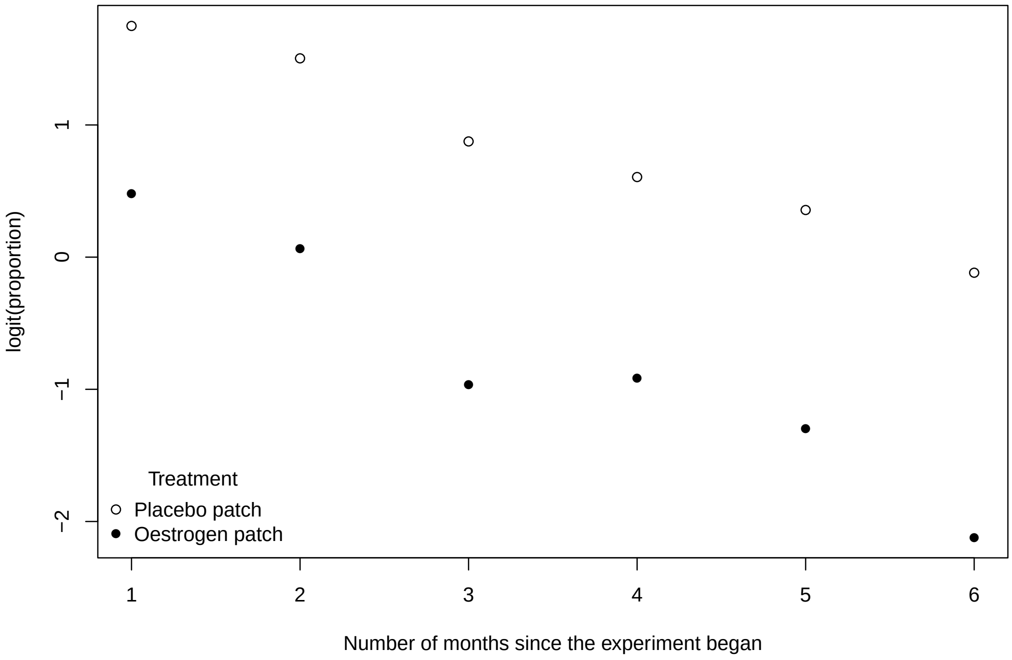

Figure 3: Evolution of the (logit of the) proportion of women with severe depression.

Several GEE with variance function $V(\mu) = \mu(1 - \mu)$, logit link function, and different structures for the working correlation matrix are fitted to the data. Then, some selection criteria available in `glmtoolbox` are used to choose the more suitable structure for the correlation matrix.

```
> data(depression)
```

```
> m1 <- glmgee(depressd ~ visit + group, id=subj, family=binomial(logit), data=depression)
> m2 <- update(m1, corstr="Exchangeable")
> m3 <- update(m1, corstr="AR-M-dependent(1)")
> m4 <- update(m1, corstr="AR-M-dependent(2)")
> m5 <- update(m1, corstr="AR-M-dependent(3)")

> a <- CIC(m1, m2, m3, m4, m5, verbose=FALSE)
> b <- QIC(m1, m2, m3, m4, m5, verbose=FALSE)
> c <- AGPC(m1, m2, m3, m4, m5, verbose=FALSE)
> d <- SGPC(m1, m2, m3, m4, m5, verbose=FALSE)
> cbind(a,QIC=b[,"QIC"],AGPC=c[,"AGPC"],SGPC=d[,"SGPC"])

  Object       Correlation  CIC    QIC     AGPC     SGPC
1     m1      Independence 7.708 383.555 304.2907 310.6233
2     m2      Exchangeable 8.048 377.815 247.9647 256.4082
3     m3 AR-M-dependent(1) 6.971 358.244 234.4696 242.9131
4     m4 AR-M-dependent(2) 7.031 363.784 234.9438 245.4982
5     m5 AR-M-dependent(3) 7.230 366.387 231.5530 244.2183
```

According to most of the selection criteria, the AR-1 structure is more suitable. Here is the summary of the GEE with that structure for the working correlation matrix:

```
> summary(m3)
Sample size
   Number of observations:  356
       Number of clusters:  61
                            Min  25%  50%  75%  Max
            Cluster sizes:    2    4    7    7    7
*********************************************************************
Model
        Variance function:  binomial
            Link function:  logit
    Correlation structure:  AR-M-dependent(1)
*********************************************************************
Coefficients
                Estimate Std.Error  z-value   Pr(>|z|)
(Intercept)      3.23604   0.51842  6.24218 4.3152e-10
visit           -0.62632   0.07477 -8.37681 < 2.22e-16
groupoestrogen -1.77723   0.54578 -3.25631  0.0011287

Dispersion     1.02842
*********************************************************************
Working correlation
     [1]   [2]   [3]   [4]   [5]   [6]   [7]
[1] 1.000 0.513 0.263 0.135 0.069 0.036 0.018
[2] 0.513 1.000 0.513 0.263 0.135 0.069 0.036
[3] 0.263 0.513 1.000 0.513 0.263 0.135 0.069
[4] 0.135 0.263 0.513 1.000 0.513 0.263 0.135
[5] 0.069 0.135 0.263 0.513 1.000 0.513 0.263
[6] 0.036 0.069 0.135 0.263 0.513 1.000 0.513
[7] 0.018 0.036 0.069 0.135 0.263 0.513 1.000
```

The Wald test and the generalized score test indicate that there is no interaction between time and the type of patch. As shown below, the p-values associated with that effect are "large".

```
> m3a <- update(m3, . ~ . + visit:group)
> anova(m3a, test="wald")
Wald test

Model 1 :  depressd ~ 1
Model 2 :  depressd ~ visit
Model 3 :  depressd ~ visit + group
Model 4 :  depressd ~ visit + group + visit:group

         Chi    df   Pr(>Chi)
```

```
1 vs 2 88.1275   1  < 2.2e-16 ***
2 vs 3 10.6036   1   0.001129 **
3 vs 4  2.2104   1   0.137082

> anova(m3a, test="score")
Generalized score test

Model 1 :  depressd ~ 1
Model 2 :  depressd ~ visit
Model 3 :  depressd ~ visit + group
Model 4 :  depressd ~ visit + group + visit:group

         Chi    df   Pr(>Chi)
1 vs 2 39.9226   1  2.642e-10 ***
2 vs 3 10.9208   1  0.0009509 ***
3 vs 4  2.3977   1  0.1215150
```

Variance estimation

Next, the estimates of $\text{Var}(\hat{\beta}_0), \text{Var}(\hat{\beta}_1), \text{Var}(\hat{\beta}_2)$ are obtained using four different estimators.

```
> cbind(model=diag(vcov(m3, type="model")),
+       robust=diag(vcov(m3, type="robust")),
+       bias.corrected=diag(vcov(m3, type="bias-corrected")),
+       jackknife=diag(vcov(m3, type="jackknife")))
                  model  robust bias.corrected jackknife
(Intercept)     0.26219 0.26875        0.29023   0.29023
visit           0.00844 0.00559        0.00583   0.00583
groupoestrogen  0.21114 0.29788        0.32441   0.32441
```

Parameter interpretation

Regardless of the type of patch (placebo or oestrogen), the probability of severe depression decreases across time. However, the odds of severe depression of women treated with oestrogen patches is approximately 83.09% $= 100 \times [1 - \exp(\hat{\beta}^{\text{group}})]$ lower than that of women treated with placebo patches, where β^{group} represents the parameter associated with the dummy variable indicating the type of patch the women were treated with.

Variable selection

As an illustration, the procedure of variable selection is applied using the strategy "hybrid forward stepwise" with the QIC as the comparison criterion. In addition, the simplest and most complex models are specified to be 1 and `1 + visit + group + visit*group`, respectively. According to this strategy, the "best" linear predictor consists of the effects of time and type of patch, but without the interaction between them. The same results are obtained in the following scenarios: (i) the "hybrid forward stepwise" is replaced by the "hybrid backward stepwise"; and (ii) the comparison criterion based on the QIC is replaced by the AGPC.

```
> stepCriterion(m3a, direction="forward", criterion="qic")

    Variance function:  binomial
        Link function:  logit
Correlation structure:  AR-M-dependent(1)
  Comparison criteria:  QIC

Initial model:
~ 1

Step 0 :
             df    QIC   QICu   AGPC   SGPC P(Chisq>)(*)
+ visit       1 413.89 409.23 273.13 279.46    < 2.2e-16
+ group       1 443.77 440.13 366.77 373.10     0.000424
<none>          465.70 463.63 385.09 389.31
```

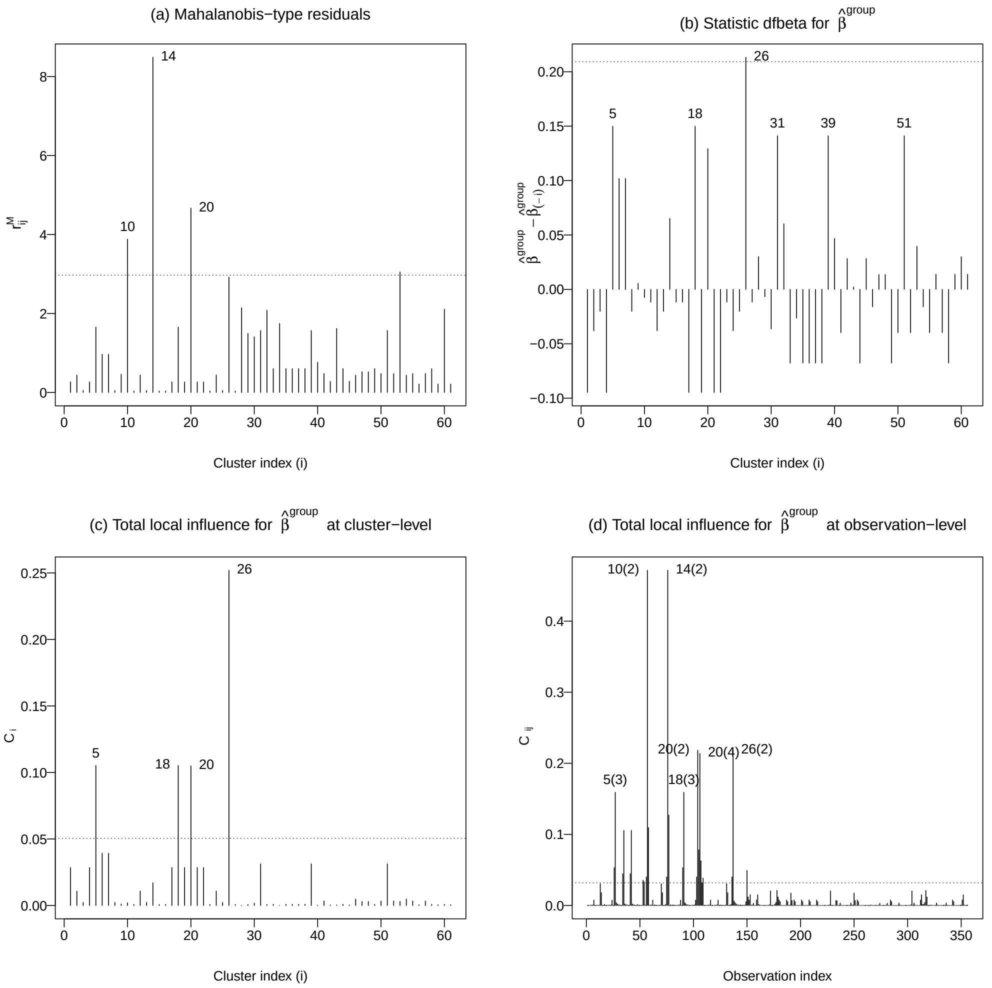

Figure 4: Some diagnostic plots for the GEE with correlation structure AR-1 fitted to the data on postnatal depression.

```
Step 1 : + visit

              df    QIC   QICu   AGPC   SGPC P(Chisq>)(*)
+ group        1 358.24 350.30 234.47 242.91     0.001129
<none>           413.89 409.23 273.13 279.46

Step 2 : + group

              df    QIC   QICu   AGPC   SGPC P(Chisq>)(*)
<none>           358.24 350.30 234.47 242.91
+ group:visit  1 358.63 351.24 234.60 245.16       0.1371
- visit        1 443.77 440.13 366.77 373.10       <2e-16

Final model:
~ visit + group
*****************************************************************************
(*) p-values of the Wald test
```

Residual analysis

```
> residuals(m3, type=”mahalanobis”, plot.it=TRUE, identify=3)
```

Mahalanobis-type residuals suggest that the women for whom the model has the lowest goodness-of-fit are those identified as 10, 14 and 20 (Figure 3(a)). Those women were treated with placebo patches and just one month after therapy began their EDPS decreased until reaching values lower than 11. However, one or two months later their EDPS values increased to 11 or higher.

Global influence

```
> dfbeta(m3, method="full", coefs="group", identify=6)
```

The plot of the dfbeta statistic for $\hat{\beta}^{\text{group}}$ at the cluster-level (Figure 3(b)) highlights the women identified as 5, 18, 26, 31, 39 and 51. The EDPS values of the women identified as 31, 39 and 51 remained higher or equal to 11 even after they were treated with oestrogen patches. Their exclusion from the dataset decreases the estimate of β^{group}, that is, increases the evidence on the effectiveness of oestrogen patches for treatment of postnatal severe depression. On the other hand, the values on the EDPS of the women identified as 5, 18 and 26 remained lower than 11 since the first or second month since therapy began, although they were treated with placebo patches, so their exclusion from the dataset also increases the evidence on the effectiveness of the oestrogen patches for treatment of postnatal severe depression.

Local influence

```
> localInfluence(m3, type="total", perturbation="cw-clusters", coefs="group",
+                plot.it=TRUE, identify=4)
> localInfluence(m3, type="total", perturbation="cw-observations", coefs="group",
+                plot.it=TRUE, identify=7)
```

According to the plot of the total local influence for $\hat{\beta}^{\text{group}}$ under the case weight perturbation scheme at the cluster-level (Figure 3(c)), the women identified as 5, 18, 20 and 26 are suspected to be influential on $\hat{\beta}^{\text{group}}$. At least 4/7 of the EDPS measurements carried out on those women were smaller than 11, although they were supplied with placebo patches. The plot of the total local influence for $\hat{\beta}^{\text{group}}$ under the case weight perturbation scheme at the observation-level (Figure 3(d)) highlights mainly two kinds of observations: (1) measurements of the EDPS in which, unlike the others measurements made on the same women, the values were lower than 11, although they were treated with placebo patches (second measurement performed on women identified as 10 and 14); (2) measurements of the EDPS performed on women treated with placebo patches and in which, for the first time for those women since the treatment began, the reported value was lower than 11, thus indicating absent severe depression, which remains until the end of the observation period (second, third and fourth measurements performed on women identified as 26, 18 and 5, respectively).

Comparison with other GEE solvers

The parameter estimates and the associated standard errors provided by the function `glmgee()` are compared with those calculated by the GEE solvers available in the packages `gee`, `geepack` and `geeM`. The results are presented in Table 4. The values obtained with the other GEE solvers are very similar to those obtained with the function `glmgee()`.

	`glmtoolbox`	`geepack`	`gee`	`geeM`
(Intercept)	3.236(0.518)	3.276(0.531)	3.214(0.514)	3.199(0.543)
visit	-0.626(0.075)	-0.630(0.077)	-0.624(0.074)	-0.633(0.077)
groupestrogen	-1.777(0.546)	-1.847(0.556)	-1.754(0.543)	-1.781(0.572)
ρ	0.51	0.48	0.47	0.51

Table 4: Parameter estimates (standard errors) of the GEE model with correlation structure AR-1 fitted to the data on severe postnatal depression.

5.4 Growth patterns of two soybean genotypes

This dataset, analyzed in Davidian and Giltinan (1995) and Pinheiro and Bates (2000) and available in the object `Soybean` of the package `nlme` (Pinheiro et al., 2022), arose from an experiment aimed at comparing growth patterns of two genotypes of soybeans: Plant Introduction (`Variety="P"`), an experimental strain, and Forrest (`Variety="F"`), a commercial variety. The average leaf weight per plant (`weight`), in grams, was measured at 14, 20, 27, 34, 41, 55, 69 and 84 days after planting (`Time`) in each plot (`Plot`). As an illustration, only plots planted in 1989 (`Year=“1989”`) are analyzed here. Table 5 describes the roles played in the analysis of the variables in the `Soybean` dataset. The graph of the data (Figure 5) indicates that the location of the response (average leaf weight per plant) increases non-linearly over time. In addition, there is an approximately proportional relationship between the mean and the standard deviation, as the variance of the response variable (in the log scale) seems to be constant. As a result, the data may be analyzed assuming that the coefficient of variation is constant, that is, using a quadratic variance function. Moreover, the graph of the data also indicates that at each time point, the location of the response is larger for the experimental strain (`Variety=“P”`) than for the commercial variety (`Variety=“F”`) of soybean.

Column	Role	Description
`Plot`	Cluster/subject identifier	Identifier of the plot
`Variety`	Explanatory variable	Treatment: “F” (experimental) or “P” (commercial)
`Year`	Explanatory variable	Year the plot was planted
`Time`	Explanatory variable	Days after planting
`weight`	Response variable	Average leaf weight per plant

Table 5: Columns in the object `Soybean` of the package `nlme`.

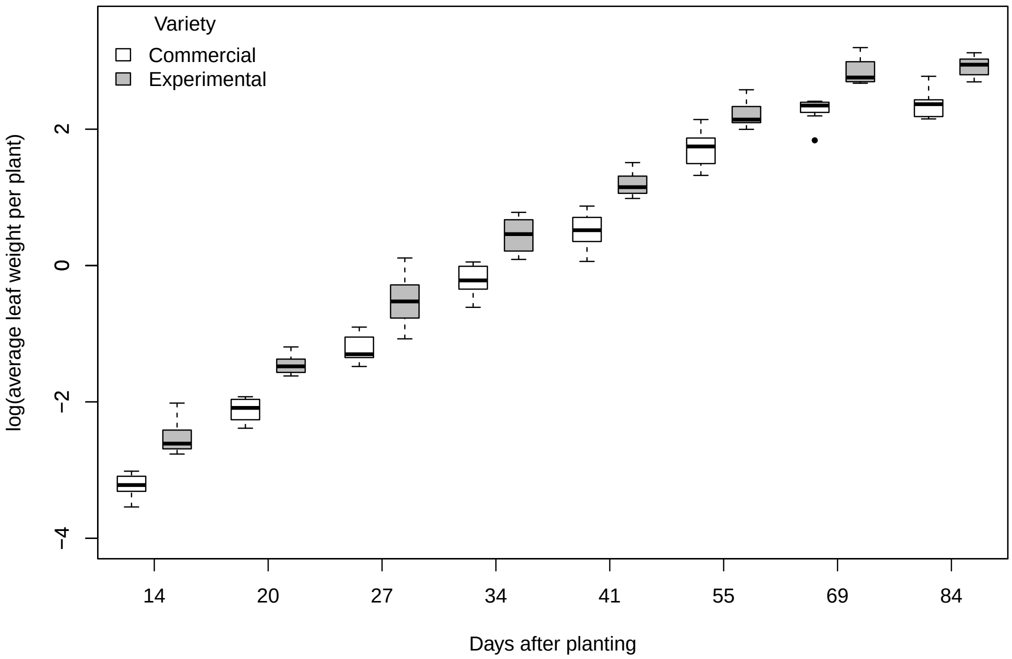

Figure 5: Average leaf weight per plant over time.

The data are analyzed using a model where the mean of the random variable Y_{ij} (jth measurement of the average leaf weight per plant performed on the ith plot) is given by the following logistic-type curve:

$$\mu_{ij} = \frac{\beta_1 + \beta_4\,\texttt{Variety}_{ij}}{1+\exp\Big(-(\texttt{Time}_{ij} - \beta_2 - \beta_5\;\texttt{Variety}_{ij})/(\beta_3 + \beta_6\,\texttt{Variety}_{ij})\Big)}, \quad i = 1,\ldots,16; j = 1,\ldots,8,$$

where $\texttt{Variety}_{ij} = 1$ if the soybean genotype is the experimental strain and $\texttt{Variety}_{ij} = 0$ otherwise. Therefore, the horizontal asymptote as $\texttt{Time} \to \infty$ (also known as the carrying capacity), the inflection point and the scale parameter of μ for commercial varieties of soybean are β_1, β_2 and β_3; and $(\beta_1 + \beta_4)$, $(\beta_2 + \beta_5)$ and $(\beta_3 + \beta_6)$ for experimental strains.

The starting value for $\boldsymbol{\beta}$ in the algorithm of parameter estimation is obtained through the built-in function `SSlogis()`.

```
> data(Soybean,package="nlme")
> Soybean2 <- subset(Soybean,Year=="1989")
> Soybean2 <- within(Soybean2,x <- ifelse(Variety=="P",1,0))
>
> m0 <- gnmgee(weight ~ SSlogis(Time,b1,b2,b3), id=Plot, family=Gamma(identity),
+              data=Soybean2)
> start <- c(coef(m0),rep(0,3))
> names(start) <- paste0("b",1:6)
> start
       b1        b2        b3        b4        b5        b6
14.185637 51.453724  7.086697  0.000000  0.000000  0.000000
```

Then, GEE models with quadratic variance function and different correlation matrix structures are fitted to the data.

```
> m1 <- gnmgee(weight ~ (b1 + b4*x)/(1 + exp(-(Time - b2 - b5*x)/(b3 + b6*x))),
+              start=start, id=Plot, family=Gamma(identity), data=Soybean2)
> m2 <- update(m1, corstr="Exchangeable")
> m3 <- update(m1, corstr="AR-M-dependent(1)")
> m4 <- update(m1, corstr="AR-M-dependent(2)")
> m5 <- update(m1, corstr="AR-M-dependent(3)")
> m6 <- update(m1, corstr="AR-M-dependent(4)")
```

As shown below, the correlation matrix structure chosen by the most of the criteria (that is, CIC, QIC, GHYC and PAC) is `AR-M-dependent(3)`.

```
> a <- CIC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> b <- QIC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> c <- GHYC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> d <- PAC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> e <- AGPC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> f <- SGPC(m1, m2, m3, m4, m5, m6, verbose=FALSE)
> cbind(a,QIC=b[,"QIC"],GHYC=c[,"GHYC"],PAC=d[,"PAC"],AGPC=e[,"AGPC"],SGPC=f[,"SGPC"])

  Object       Correlation   CIC      QIC  GHYC    PAC    AGPC    SGPC
1     m1      Independence 6.951 6163.648 8.126 0.9847 90.5844 95.2200
2     m2      Exchangeable 6.951 6163.648 7.552 0.9785 86.8152 92.2233
3     m3 AR-M-dependent(1) 6.795 6098.876 6.640 0.9753 86.1055 91.5136
4     m4 AR-M-dependent(2) 6.713 6095.808 6.622 0.9737 87.7812 93.9619
5     m5 AR-M-dependent(3) 6.708 6094.956 6.621 0.9736 89.7920 96.7453
6     m6 AR-M-dependent(4) 6.752 6115.573 6.673 0.9741 91.3912 99.1171
```

The chosen model is summarized as follows:

```
> summary(m5)

Sample size
   Number of observations:  128
       Number of clusters:  16
             Cluster size:  8
*************************************************************
Model
        Variance function:  Gamma
            Link function:  identity
    Correlation structure:  AR-M-dependent(3)
*************************************************************
Coefficients
           Estimate Std.Error  z-value Pr(>|z|)
b1         10.58794   0.54866 19.29779  < 2e-16
b2         52.08512   0.99860 52.15828  < 2e-16
b3          7.01786   0.19565 35.87033  < 2e-16
b4          7.48960   0.88795  8.43475  < 2e-16
b5         -0.77453   1.29528 -0.59797  0.54986
b6          0.09913   0.24511  0.40441  0.68591

Dispersion  0.05686
```

```
************************************************************
Working correlation
      [1]   [2]   [3]   [4]   [5]   [6]   [7]   [8]
[1] 1.000 0.253 0.151 0.053 0.025 0.010 0.004 0.002
[2] 0.253 1.000 0.253 0.151 0.053 0.025 0.010 0.004
[3] 0.151 0.253 1.000 0.253 0.151 0.053 0.025 0.010
[4] 0.053 0.151 0.253 1.000 0.253 0.151 0.053 0.025
[5] 0.025 0.053 0.151 0.253 1.000 0.253 0.151 0.053
[6] 0.010 0.025 0.053 0.151 0.253 1.000 0.253 0.151
[7] 0.004 0.010 0.025 0.053 0.151 0.253 1.000 0.253
[8] 0.002 0.004 0.010 0.025 0.053 0.151 0.253 1.000
```

These results suggest that only the horizontal asymptote as `Time` $\rightarrow \infty$ depends on the soybean variety. Their estimates are 10.588 grams and 18.078 grams for commercial varieties and experimental soybean strains, respectively.

5.5 Amenorrhea rates over time

The dataset of this example, available in the object `amenorrhea` of `glmtoolbox` and comprised of the columns named `ID`, `Dose`, `Time`, and `amenorrhea` (see Table 6), arose from a longitudinal clinical trial of contracepting women (Machin et al., 1988; Fitzmaurice et al., 2011). A total of 1151 women completed menstrual diaries. The diary data were used to generate a binary sequence for each woman, indicating whether she had experienced amenorrhea (the absence of menstrual bleeding for a specified number of days) on the day of randomization and three additional 90-day intervals. This trial compared the two treatments (injections of 100 mg or 150 mg of depot-medroxyprogesterone acetate (DMPA)) in terms of how amenorrhea rates change over time with continued use of the contraceptive method. Figure 6 shows that amenorrhea rates increase across treatments, but that it appears that women treated with 150 mg of DMPA are more likely to experience amenorrhea than those treated with 100 mg of DMPA at each time point. Moreover, Figure 6 shows that the proportion of women experiencing amenorrhea (on the logit scale) increases non-linearly over time. A feature of this clinical trial is that there was substantial dropout. This is when a woman skips a particular injection and never returns for subsequent injections. Indeed, 38% of the women dropped out before the trial ended; 17.2% dropped out after receiving only one injection of DMPA, 13.5% dropped out after receiving only two injections of DMPA, and 7.3% dropped out after receiving three injections of DMPA. The subsequent statistical analysis is performed using the weighted GEE method, as it is assumed that the missing data pattern is better described by MAR than MCAR.

Column	Role	Description
`ID`	Cluster/subject identifier	Identifier of the woman
`Dose`	Explanatory variable	Treatment: “100mg” or “150mg” of DMPA
`Time`	Explanatory variable	Number of 90-day intervals since the trial began
`amenorrhea`	Response variable	1 if experienced amenorrhea; 0 otherwise

Table 6: Columns in the object `amenorrhea` of the package `glmtoolbox`.

The data are analyzed using a model in which the probability of the ith woman experienced amenorrhea at time j, denoted here by μ_{ij}, is such that

$$\text{logit}(\mu_{ij}) = 1 + \texttt{Time} + \texttt{Time}^2 + \texttt{Dose}.$$

For the missingness model the following systematic component is considered

$$\text{logit}(\pi_{ij}) = 1 + \texttt{CTime} + \texttt{Dose} + \texttt{ylag1},$$

where `CTime` is a categorical version of the explanatory variable `Time` and `ylag1` is defined to be $y_{i,j-1}$ if $j > 1$ and 0 if $j = 1$. In addition, the structure of the working correlation matrix is set to be `AR-M-dependent(1)`. The observation-specified weighted GEE method results are the following:

```
> data(amenorrhea)
> amenorrhea2 <- within(amenorrhea,{Ctime <- factor(Time)
+                                   Ctime <- relevel(Ctime,ref="1")
+                                   ylag1 <- c(0,amenorrhea[-length(ID)])
+                                   ylag1 <- ifelse(Time==0,0,ylag1)})
>
```

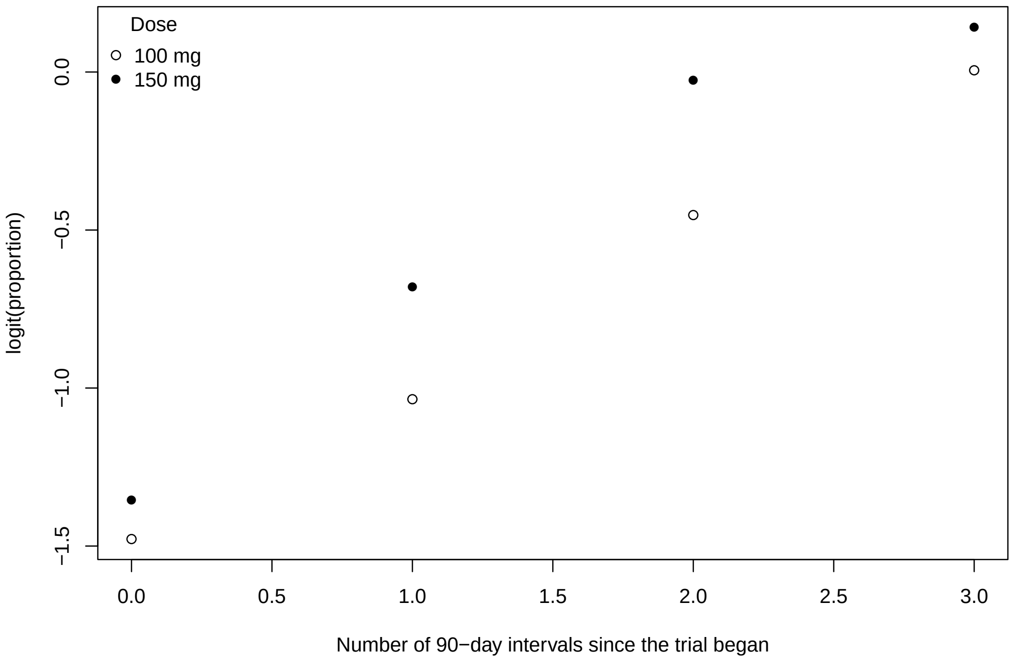

Figure 6: Amenorrhea rates over time.

```
> fit1 <- wglmgee(amenorrhea ~ poly(Time,2) + Dose | Ctime + Dose + ylag1,
+                 family=binomial, data=amenorrhea2, id=ID, corstr="AR-M-dependent(1)",
+                 scale.fix=TRUE, scale.value=1, level="observations")
> summary(fit1)

Clusters by dropout time
 Time 1    2    3    4   Freq   %
   ---- ---- ---- ---- | ---- ----
      X    .    .    . |  198 17.2
      X    X    .    . |  155 13.5
      X    X    X    . |   84  7.3
      X    X    X    X |  714   62
   ---- ---- ---- ---- | ---- ----
                       | 1151  100
*************************************************************
Coefficients of missingness model
            Estimate Std.Error z-value  Pr(>|z|)
(Intercept)   2.4349    0.1401 17.3845 < 2.2e-16 ***
Ctime1       -0.7247    0.1438 -5.0399 4.659e-07 ***
Ctime2       -0.5911    0.1469 -4.0250 5.698e-05 ***
Dose150mg    -0.0174    0.1049 -0.1663    0.8679
ylag1        -0.5765    0.1122 -5.1369 2.793e-07 ***
*************************************************************
Observation-specific Weighted GEE

        Variance function:  binomial
            Link function:  logit
    Correlation structure:  AR-M-dependent(1)
*************************************************************
Coefficients
               Estimate Std.Error z-value Pr(>|z|)
(Intercept)     -0.6835    0.0750 -9.1136  < 2e-16 ***
poly(Time, 2)1  40.7447    2.2598 18.0301  < 2e-16 ***
poly(Time, 2)2  -4.6883    1.9528 -2.4008  0.01636 *
```

```
Dose150mg        0.2437    0.1061  2.2975  0.02159 *

Dispersion       1.0000
*************************************************************
Working correlation
     [1]   [2]   [3]   [4]
[1] 1.000 0.414 0.171 0.071
[2] 0.414 1.000 0.414 0.171
[3] 0.171 0.414 1.000 0.414
[4] 0.071 0.171 0.414 1.000
```

According to the missingness model, the probability of remaining in the trial increases over time, regardless of the amenorrhea status reported by the women in the previous measurement. In addition, the missingness model indicates that the probability of remaining in the trial is higher for women without amenorrhea in their previous measurement, regardless of how many DMPA injections they have received. Therefore, the MAR assumption seems more appropriate than MCAR. Moreover, according to the model for μ, the odds of experiencing amenorrhea is approximately $27.6\% = 100 \times [\exp(\hat{\beta}^{\text{Dose}}) - 1]$ higher in women treated with 150 mg of DMPA than those treated with 100 mg of DMPA, where β^{Dose} represents the parameter associated with the explanatory variable `Dose`. This is irrespective of the number of 90-day intervals since the experiment began. The results of the cluster-specific weighted GEE approach (do not show here) are very similar.

Acknowledgments

The authors are grateful to the Editor and the referees for their helpful comments and suggestions that have led to significant improvements of this paper.

References

D. Boos. On generalized score tests. *The American Statistician* , 46(4):327–33, 1992. [p110]

V. Carey. *gee: Generalized Estimation Equation Solver*, 2022. URL `https://CRAN.R-project.org/package=gee`. R package version 4.13-23. [p105]

V. Carey and Y.-G. Wang. Working covariance model selection for generalized estimating equations. *Statistics in Medicine*, 30(26):3117–3124, 2011. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4300`. [p111]

R. Cook. Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):133–169, 1986. [p112]

M. Davidian and D. M. Giltinan. *Nonlinear models for repeated measurement data*, volume 62. CRC press, 1995. [p126]

P. Diggle, P. J. Diggle, P. Heagerty, K.-Y. Liang, S. Zeger, et al. *Analysis of longitudinal data*. Oxford university press, 2002. [p116]

G. Fitzmaurice, N. Laird, and J. Ware. *Applied Longitudinal Analysis. 2nd ed.* John Wiley & Sons, 2011. [p115, 129]

L. Fu, Y. Hao, and Y.-G. Wang. Working correlation structure selection in generalized estimating equations. *Computational Statistics*, 33:983–96, 2018. [p111]

M. Gosho. Criteria to select a working correlation structure in sas. *Journal of Statistical Software, Code Snippets*, 57(1):1–10, 2014. doi: 10.18637/jss.v057.c01. URL `https://www.jstatsoft.org/index.php/jss/article/view/v057c01`. [p110]

M. Gosho, C. Hamada, and I. Yoshimura. Criterion for the selection of a working correlation structure in the generalized estimating equation approach for longitudinal balanced data. *Communications in Statistics - Theory and Methods*, 40(21):3839–3856, 2011. [p110]

A. Gregoire, R. Kumar, B. Everitt, and J. Studd. Transdermal oestrogen for treatment of severe postnatal depression. *The Lancet*, 347(9006):930–933, 1996. [p121]

B. Hammill and J. Preisser. A sas/iml software program for gee and regression diagnostics. *Computational Statistics & Data Analysis*, 51(2):1197–1212, 2006. [p111, 112, 114]

L. Hin, V. Carey, and Y.-G. Wang. Criteria for working-correlation-structure selection in gee: Assessment via simulation. *The American Statistician*, 61(4):360–364, 2007. URL `http://www.jstor.org/stable/27643940`. [p111]

L.-Y. Hin and Y.-G. Wang. Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine*, 28(4):642–658, 2009. [p110]

S. Højsgaard, U. Halekoh, and J. Yan. The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2):1–11, 2005. [p105]

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013. [p113]

K.-M. Jung. Local influence in generalized estimating equations. *Scandinavian Journal of Statistics*, 35 (2):286–294, 2008. [p112]

N. Laird. Missing data in longitudinal studies. *Statistics in medicine*, 7(1-2):305–315, 1988. [p109]

K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73: 13–22, 1986. [p105, 106, 109]

S. Lipsitz and G. Fitzmaurice. *Generalized estimating equations for longitudinal data analysis*, chapter chapter3. CRC Press, 2008. [p105, 109]

S. Lipsitz, N. Laird, and D. Harrington. Using the jackknife to estimate the variance of regression estimators from repeated measures studies. *Communications in Statistics - Theory and Methods*, 19(3): 821–845, 1990. [p109]

D. Machin, T. Farley, B. Busca, M. Campbell, and C. d'Arcangues. Assessing changes in vaginal bleeding patterns in contracepting women. *Contraception*, 38:165–179, 1988. [p129]

M. Maechler, P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, and M. Anna di Palma. *robustbase: Basic Robust Statistics*, 2022. R package version 0.95-0. [p116]

L. Mancl and T. DeRouen. A covariance estimator for gee with improved small-sample properties. *Biometrics*, 57(1):126–134, 2001. doi: doi:10.1111/j.0006-341x.2001.00126.x. [p109]

P. McCullagh and J. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989. URL `http://books.google.com/books?id=h9kFH2_FfBkC`. [p105, 110]

L. McDaniel, N. Henderson, and P. Rathouz. Fast pure R implementation of GEE: application of the Matrix package. *The R Journal*, 5:181–187, 2013. URL `https://journal.r-project.org/archive/2013-1/mcdaniel-henderson-rathouz.pdf`. [p105]

W. Pan. Akaike's information criterion in generalized estimating equations. *Biometrics*, 57(1):120–125, 2001. URL `http://www.jstor.org/stable/2676849`. [p110, 113]

M. C. Pardo and R. Alonso. Working correlation structure selection in gee analysis. *Statistical Papers*, 60(5):1447–1467, 2019. URL `https://doi.org/10.1007/s00362-017-0881-0`. [p111]

J. Pinheiro and D. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer New York, 2000. ISBN 9780387989570. URL `https://books.google.com.co/books?id=N3WeyHFbHLQC`. [p126]

J. Pinheiro, D. Bates, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2022. URL `https://CRAN.R-project.org/package=nlme`. R package version 3.1-160. [p126]

J. Preisser and B. Qaqish. Deletion diagnostics for generalised estimating equations. *Biometrika*, 83(3): 551–562, 1996. [p111, 112, 114]

J. Preisser, K. Lohman, and P. Rathouz. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine*, 21:3035–3054, 2002. [p115, 116]

J. Robins, A. Rotnitzky, and L. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:122–129, 1995. [p115, 116]

A. Rotnitzky and N. Jewell. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77(3):485–497, 1990. URL `http://www.jstor.org/stable/2336986`. [p110]

R. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, 61(3):439–447, 1974. [p105]

J. Xu, J. Zhang, and L. Fu. Variable selection in generalized estimating equations via empirical likelihood and gaussian pseudo-likelihood. *Communications in Statistics - Simulation and Computation*, 48(4):1239–1250, 2019. [p113]

J. Yan. geepack: Yet another package for generalized estimating equations. *R-News*, 2/3:12–14, 2002. [p105]

A. Zeileis and Y. Croissant. Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, 34(1):1–13, 2010. doi: 10.18637/jss.v034.i01. [p116]

X. Zhu and Z. Zhu. Comparison of criteria to select working correlation matrix in generalized estimating equations. *Chinese Journal of applied probability and statistics*, 5:515–30, 2013. [p111]

L.H. Vanegas
Departamento de Estadística, Universidad Nacional de Colombia
Ciudad Universitaria, Bogotá
Colombia
`lhvanegasp@unal.edu.co`

L.M. Rondón
Departamento de Estadística, Universidad Nacional de Colombia
Ciudad Universitaria, Bogotá
Colombia
`lmrondonp@unal.edu.co`

G.A. Paula
Instituto de Matemática e Estatística, Universidade de São Paulo
Rua do Matão, 1010, São Paulo
Brazil
`giapaula@ime.usp.br`